PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :
C12N 15/00, 7/00, C12Q 1/70
G01N 33/569, C07K 13/00, 15/00
C12N 1/20, 5/00

A1

(11) International Publication Number: WO 87/07906

(43) International Publication Date: 30 December 1987 (30.12.87)

(21) International Application Number: PCT/EP87/00326

(22) International Filing Date: 22 June 1987 (22.06.87)

(31) Priority Application Number: 86401380.0 (EP)

(32) Priority Date: 23 June 1986 (23.06.86)

(33) Priority Countries: GB, et al.

(71) Applicant: INSTITUT PASTEUR [FR/FR]; 25-28, rue du Dr. Roux, F-75015 Paris (FR).

(72) Inventors: ALIZON, Marc ; 71, rue du Cardinal Lemoine, F-75005 Paris (FR). SONIGO, Pierre ; 23, rue Gutenberg, F-75015 Paris (FR). WAIN-HOBSON, Simon ; 3, rue Jean de la Fontaine, F-78180 Montigny-les-Bretonneux (FR). MONTAGNIER, Luc ; 21, rue de Malabry, F-92350 Le Plessis Robinson (FR).

(74) Agents: GUTMANN, Ernest et al.; S.C. Ernest Gutmann - Yves Plasseraud, 67, bld. Haussmann, F-75008 Paris (FR).

(81) Designated States: AU, DK, JP.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: VARIANTS OF LAV VIRUSES, THEIR DNA- AND PROTEIN-COMPONENTS AND THEIR USES, PARTICULARLY FOR DIAGNOSTIC PURPOSES AND FOR THE PREPARARTION OF IMMUNOGENIC COMPOSITIONS

(57) Abstract

Two variants of LAV viruses capable of causing acquired immunosuppressive syndrome (AIDS), which virus variants have been designated as LAV_{ELI} and LAV_{MAL}. Their DNAs and antigens can be used for the diagnostic of AIDs or pre-AIDS.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
DE	Germany, Federal Republic of	LU	Luxembourg	TG	Togo
DK	Denmark	MC	Monaco	US	United States of America
FI	Finland	MG	Madagascar		

VARIANTS OF LAV VIRUSES, THEIR DNA- AND PROTEIN-COMPONENTS AND THEIR USES, PARTICULARLY FOR DIAGNOSTIC PURPOSES AND FOR THE PREPARATION OF IMMUNOGENIC COMPOSITIONS

The present invention relates to viruses capable of inducing lymphadenopathies (denoted below by the abbreviation LAS) acquired immuno-depressive syndromes (denoted below by the abbreviation AIDS), to antigens of said viruses, particularly in a purified form, and to processes for producing these antigens, particularly antigens of the envelopes of these viruses. The invention also relates to polypeptides, whether glycosylated or not, encoded by said DNA sequences.

The invention also relates to cloned DNA sequences hybridizable to genomic RNA and DNA of the new lymphadenopathy associated viruses (LAV) disclosed hereafter, to processes for their preparation and their uses. It relates more particularly to stable probes including a DNA sequence which can be used for the detection of the new LAV viruses or related viruses or DNA proviruses in any medium, particularly biological, samples, containing of any them.

An important genetic polymorphism has been recognized for the human retrovirus at the origin of the acquired immune deficiency syndrome (AIDS) and other diseases, like lymphadenopathy syndrome (LAS), AIDS-related complex (ARC) and probably some encephalopathies (for review see Weiss, 1984). Indeed all of the isolates analyzed until now have a distinct restriction map, even if recovered from the same place and time (BENN et al., 1985). Identical restriction maps have only been observed for the first two isolates designated lymphadenopathy-associated virus, LAV (ALIZON et al., 1984) and human T-cell lymphotropic virus type 3, HTLV-3 (HAHN et al., 1984) and thus appears as an exception.

The genetic polymorphism of the AIDS virus was better assessed after the determination of the complete nucleotide sequence of LAV (WAIN-HOBSON et al., 1985), HTLV-3 (RATNER et al., 1985 ; MUESING et al., 1985) and of a third isolate designated AIDS-associated retrovirus, ARV (SANCHEZ-PESCADOR et al., 1985). In particular it appeared that, besides the nucleic acid variations responsible for the restriction map polymorphism, isolates could differ significantly at the protein level, especially in the envelope (up to 13 % of difference between ARV and LAV), by both amino-acids substitutions and reciprocal insertions-deletions (RABSON and MARTIN, 1985).

Nevertheless the differences mentioned above do not go as far as to destroy a level of immunological relationship sufficient, as evidenced by the capabilities of similar proteins, i. e. core proteins of similar nature, such as the p25 proteins, or of similar envelope glycoproteins, such as the 110-120 kD glycoproteins, to immunologically cross-react. Accordingly the proteins of any of said LAV viruses can be used for the <u>in vitro</u> detection of antibodies induced <u>in vivo</u> and present in biological fluids obtained from individuals infected with the other LAV variants. Therefore these viruses are grouped in a class of LAV viruses, hereafter generally said to belong to the class of LAV-1 viruses.

The invention stems from the discovery of new viruses which although held as responsible of diseases which are clinically related to AIDS and still belonging to the class of "LAV-1 viruses", differ genetically to a much larger extent from the above mentioned LAV variants.

The new viruses are basically characterized by the DNA sequences which are shown in Figures 7A to 7J (LAV_{ELI}) and figures 8A to 8I ($LAVM_{AL}$) respectively.

The invention further relates to variants of

the new viruses the RNAs of which or the related cDNAs derived from said RNAs are hybridizable to corresponding parts of the cDNAs of either LAV_{ELI} or LAV_{MAL}.

The invention also relates to the DNAs themselves of said viruses, including DNA fragments derived therefrom hybridizable with the genomic RNA of either LAV_{ELI} or LAV_{MAL}. Particularly said DNAs consist of said cDNAs or cDNA fragments or of recombinant DNAs containing said cDNAs or cDNA fragments.

It further relates to DNA recombinants containing DNAs or cDNA fragments of either LAV_{ELI} or LAV_{MAL} or of related viruses. It is of course understood that fragments which would include some deletions or mutations which would not substantially alter their capability of also hybridizing with the retroviral genomes of LAV_{ELI} or LAV_{MAL} are to be considered as forming obvious equivalents of the DNAs or DNA fragments more specifically referred to hereabove.

The invention also relates more specifically to cloned probes which can be made starting from any DNA fragment according to the invention, thus to recombinant DNAs containing such fragments, particularly any plasmids amplifiable in procaryotic or eucaryotic cells and carrying said fragments.

Using the cloned DNA containing a DNA fragment of LAV_{ELI} or of LAV_{MAL} as a molecular hybridization probe - either by marking with radionucleotides or with fluorescent reagents - LAV virion RNA may be detected directly e. g. in the blood, body fluids and blood products (e.g. of the antihemophylic factors such as Factor VIII concentrates). A suitable method for achieving that detection comprises immobilizing virus onto said a support e.g. nitrocellulose filters, etc., disrupting the virion and hybridizing with labelled (radiolabelled or "cold" fluorescent- or enzyme-labelled) probes. Such

an approach has already been developed for Hepatitis B virus in peripheral blood (according to SCOTTO J. et al. Hepatology (1983), 3, 379-384).

Probes according to the invention can also be used for rapid screening of genomic DNA derived from the tissue of patients with LAV related symptoms, to see if the proviral DNA or RNA present in host tissue and other tissues are related to LAV_{ELI} or LAV_{MAL}.

A method which can be used for such screening comprise the following steps : extraction of DNA from tissue, restriction enzyme cleavage of said DNA, electrophoresis of the fragments and Southern blotting of genomic DNA from tissues, subsequent hybridization with labelled cloned LAV provival DNA. Hybridization in situ can also be used.

Lymphatic fluids and tissues and other non-lymphatic tissues of humans, primates and other mammalian species can also be screened to see if other evolutionnary related retrovirus exist. The methods referred to hereabove can be used, although hybridization and washings would be done under non stringent conditions.

The DNA according to the invention can be used also for achieving the expression of LAV viral antigens for diagnostic purposes as well as far the production of a vaccine against LAV. Fragments of particular advantage in that respect will be discussed later.

The methods which can be used are multifold :

a) DNA can be transfected into mammalian cells with appropriate selection markers by a variety of techniques, calcium phosphate precipitation, polyethylene glycol, protoplast-fusion, etc..

b) DNA fragments corresponding to genes can be cloned into expression vectors for *E. coli* , yeast- or mammalian cells and the resultant proteins purified.

c) The provival DNA can be "shot-gunned" (fragmented) into procaryotic expression vectors to generate fusion polypeptides. Recombinant producing antigenically competent fusion proteins can be identified by simply screening the recombinants with antibodies against LAV_{ELI} or LAV_{MAL} antigens.

Particular reference in that respect is made to those portions of the genomas of LAV_{ELI} and LAV_{MAL} which, in the drawings, are shown to belong to open reading frames and which encode the products having the polypeptidic backbones shown.

More particularly, the invention relates to the different polypeptides which appear in figures 7A to 8I. Methods disclosed in European application 0 178 978 and in PCT application PCT/EP 85/00548 filed on Oct.18, 1985 are applicable for the production of such peptides from the corresponding viruses.

The present invention further aims at providing polypeptides containing sequences in common with polypeptides comprising antigenic determinants included in the proteins encoded and expressed by the LAV_{ELI} or of LAV_{MAL} genome. An additional object of the invention is to further provide means for the detection of proteins related to these LAV viruses, particularly for the diagnosis of AIDS or pre-AIDS or, to the contrary, for the detection of antibodies against the LAV virus or proteins related therewith, particularly in patients afflicted with AIDS or pre-AIDS or more generally in asymtomatic carriers and in blood-related products. Finally the invention also aims at providing immunogenic polypeptides, and more particularly protective polypeptides for use in the preparation of vaccine compositions against AIDS or related syndroms.

The invention relates also to polypeptide fragments having lower molecular weights and having

peptide sequences or fragments in common with those shown in figures 7A to 8I. Fragments of smaller sizes may be obtained by resorting to known techniques. For instance such a method comprises cleaving the original larger polypeptide by enzymes capable of cleaving it at specific sites. By way of examples of such proteins, may be mentioned the enzyme of Staphylococcyus aureus V8, α-chymotrypsine, "mouse sub-maxillary gland protease" marketed by the BOEHRINGER company, Vibrio alginolyticus chemovar iophagus collagenase, which specifically recognizes said peptides Gly-Pro and Gly-Ala, etc.

Other features of this invention will appear in the following disclosure of the data obtained starting from LAV_{ELI} and LAV_{MAL}, in relation to the drawings in which :

- Figs 1A and 1B provide restriction maps of the genomas of LAV_{ELI} and LAV_{MAL} as compared to LAV_{BRU} (a known LAV isolate deposited at CNCM under number I-232 on July 15th, 1983) ;
- Fig. 2 shows the comparative maps setting forth the relative positions of the open reading frames of the above genomas ;
- Figs. 3A-3F (sometimes also designated globally hereafter by fig. 3) indicate the relative correspondance between the proteins (or glycoproteins) encoded by the open reading frames, whereby aminoacid residues of protein sequences of LAV_{ELI} and LAV_{MAL} are in vertical alinment with corresponding aminoacid residues (numbered) of corresponding or homologous proteins or glycoproteins of LAV_{BRU} ;
- Figs. 4A-4B (sometimes also designated globally hereafter by fig. 4) provide for quantitation of the sequence divergence between homologous proteins of LAV_{BRU}, LAV_{ELI} and LAV_{MAL} ;
- Fig. 5 shows diagrammatically the degree of divergence of the different virus enveloppe proteins ;
- Figs. 6A and 6B (or Fig. 6 when viewed altogether) render apparent the direct repeats which appear in the

proteins of the different AIDS virus isolates.

- Figs. 7A-7J and 8A-8I show the full nucleotidic sequences of LAV_{ELI} and LAV_{MAL} respectively.

RESULTS

Characterization and molecular cloning of two African isolates.

The different AIDS virus isolates concerned are designated by three letters of the patients name, LAV_{BRU} refeRring to the prototype AIDS virus isolated in 1983 from a French homosexual patient with LAS and thought to have been infected in USA in the preceding years (Barré-Sinoussi et al., 1983). Both of the African patients originated from Zaire ; LAV_{ELI} was recovered in 1983 from a 24 year old woman with AIDS, and LAV_{MAL} in 1985 from a 7 year old boy with ARC, probably infected in 1981 after a blood-transfusion in Zaire, since his parents were LAV-seronegative.

Recovery and purification of each of the two viruses were performed according to the method disclosed in European Patent Aplication 84 401834/138 667 filed on September 9, 1984.

LAV_{ELI} and LAV_{MAL} are indistinguishable from the previously characterized isolates by their structural and biological properties in vitro. Virus metabolic labelling and immune precipitation by patients ELI and MAL sera, as well as reference sera, showed that the proteins of LAV_{ELI} and LAV_{MAL} had the same molecular weight (MW) and cross-reacted immunologically with those of prototype AIDS virus (data not shown) of the "LAV 1" class.

Reference is again made to European Application 178 978 and International Application PCT/EP 85/00548 as concerns the purification, mapping and sequencing procedures used herein. See also "experimental procedures" and "legends of the figures" hereafter.

Primary restriction enzyme analysis of LAV_{ELI} and LAV_{MAL} genomes was done by southern blot with total DNA derived from acutely infected lymphocytes, using cloned LAV_{BRU} complete genome as probe. Overall cross-hybridization was observed under stringent conditions, but the restriction profiles of the Zairian isolates were clearly different. Phage lambda clones carrying the complete viral genetic information were obtained and further characterized by restriction mapping and nucleotide sequence analysis ; clone E-H12 is derived from LAV_{ELI} infected cells and contains an integrated provirus with 5' flanking cellular sequences but a truncated 3' long terminal repeat (LTR) ; clone M-H 11 was obtained by complete HindIII restriction of DNA from LAV_{MAL}-infected cells, taking advantage of the existence of a unique HindIII site in the LTR. M-H 11 is thus probably derived from unintegrated viral DNA since that species was at least ten times more abundant than integrated provirus.

Figure 1B gives a comparaison of the restriction maps of LAV_{ELI}, LAV_{MAL} and prototype LAV_{BRU}, all three being derived from their nucleotide sequences, as well of three Zairian isolates previously mapped for seven restriction enzymes (Benn et al., 1985). Despite this limited number, all of the profiles are clearly different (out of the 23 sites making up the map of LAV^{BRU} only seven are present in all six maps presented), confirming the genetic polymorphism of the AIDS virus. No obvious relationship is apparent between the five Zairian maps, and all of their common sites are also found in LAV_{BRU}.

Conservation of the genetic organization.

The genetic organization of LAV_{ELI} and LAV_{MAL} as deduced from the complete nucleotide sequences of their cloned genomes is identical to that found in other

isolates, i.e. 5'gag-pol-central region-env-F3'. Most noticeable is the conservation of the "central region" (fig. 2), located between the pol and env genes, which is composed of a series of overlapping open reading frames (orf) we had previously designated Q, R, S, T, and U after observing a similar organization in the ovine lentivirus visna (Sonigo et al., 1985). The product of orf S (also designated "tat") is implicated in the transactivation of virus expression (Sodroski et al., 1985 ; Arya et al., 1985) ; the biological role of the product of orf Q (also designated "sor" or orf A) is still unknown (Lee et al., 1986 ; kang et al., 1986). Of the three other orfs (R, T, and U), only orf R is likely to be a seventh viral gene, for the following reasons : the exact conservation of its relative position with respect to Q and S (fig. 2), the constant presence of a possible splice acceptor and of a consensus AUG initiator codon, its similar codon usage with respect to viral genes, and finally the fact that the variation of its protein sequence within the different isolates is comparable to that of gag, pol and Q (see fig. 4).

Also conserved are the sizes of the U3, R and U5 elements of the LTR (data not shown), the location and sequence of their regulatory elements such as TATA box and AATAAA polyadenylation signal, and their flanking sequences i.e. primer binding site (PBS) complementary to 3' end of $tRNA^{LYS}_{3}$ and polypurine tract (PPT). Most of the genetic variability within the LTR is located in the 5' half of U3 (which encodes a part of orf F) while the 3' end of U3 and R, which carry most of the cis-acting regulatory elements : promoter, enhancer and trans-activating factor receptor (Rosen et al., 1985), as well as the U5 element are well-conserved.

Overall, it clearly appears that the Zairian

isolates belong to the same type of retrovirus as the previously sequenced isolates of American or European origin.

Variability of the viral proteins.

Despite their identical genetic organization, these isolates show substantial differences in the primary structure of their proteins. The amino acid sequences of LAV_{ELI} and LAV_{MAL} proteins are presented in figures 3A-3F (to be examined in conjunction with Figs. 7A-7J and 8A-8I), aligned with those of LAV_{BRU} and ARV 2. Their divergence was quantified as the percentage of amino-acids substitutions in two-by-two alignments (Fig. 4). We have also scored the number of insertions and deletions that had to be introduced in each of these alignments.

Three general observations can be made. First, the protein sequences of the African isolates are more divergent from LAV_{BRU} than are those of HTLV-3 and ARV 2 (Fig. 4A) ; similar results are obtained if ARV 2 is taken as reference (not shown). The range of genetic polymorphism between isolates of the AIDS virus is considerably greater than previously observed. Second, our two sequences confirm that the envelope is more variable than the gag and pol genes. Here again, the relatively small difference observed between the env of LAV_{BRU} and HTLV-3 appears as an exception. Third, the mutual divergence of the two African isolates (Fig. 4B) is comparable to that between LAV_{BRU} and either of them; as far as we can extrapolate from only three sequenced isolates from the USA and Europe and two from Africa, this is indicative of a wider evolution of the AIDS virus in Africa.

<u>gag and pol</u> : Their greater degree of conservation compared to the envelope is consistent with their encoding important structural or enzymatic activities.

Of the three mature gag proteins, the p25 which was the first recognized immunogenic protein of LAV (Barré-Sinoussi et al., 1983) is also the better conserved (fig. 3). In gag and pol, differences between isolates are principally due to point mutations, and only a small number of insertional or deletional events is observed. Among these, we must note the presence in the overlapping part of gag and pol of LAV_{BRU} of an insertion of 12 aminoacids (AA) which is encoded by the second copy of a 36 bp direct repeat present only in this isolate and in HTLV-3. This duplication was omitted because of a computing error in the published sequence of LAV_{BRU} (position 1712, Wain-Hobson et al., 1985) but was indeed present in the HTLV-3 sequences (Ratner et al., 1985 ; Muesing et al., 1985).

env : Three segments can be distinguished in the envelope glycoprotein precursor (Allan et al., 1985 ; Montagnier et al., 1985 ; DiMarzoVeronese et al., 1985). The first is the signal peptide (positions 1-33 in fig. 3), and its sequence appears as variable ; the second segment (pos. 34-530) forms the outer membrane protein (OMP or gp110) and carries most of the genetic variations, and in particular almost all of the numerous reciprocal insertions and deletions ; the third segment (531-877) is separated from the OMP by a potential cleavage site following a constant basic stretch (Arg-Glu-Lys-Arg) and forms the transmembrane protein (TMP or gp 41) responsible for the anchorage of the envelope glycoprotein in the cellular membrane. A better conservation of the TMP than the OMP has also been observed between the different murine leukemia viruses (MLV, Koch et al., 1983), and could be due to structural constraints.

From the alignment of figure 3 and the graphical representation of the envelope variability

shown in figure 5, we clearly see the existence of conserved domains, with little or no genetic variation, and hypervariable domains, in which even the alignment of the different sequences is very difficult, because of the existence of a large number of mutations and of reciprocal insertions and deletions. We have not included the sequence of the envelope of the HTLV-3 isolate since it so close to that of LAV_{BRU} (cf. fig. 4), even in the hypervariable domains, that it did not add anything to the analysis. While this graphical representation will be refined by more sequence data, the general profile is already apparent, with three hypervariable domains (Hyl, 2 and 3) all being located in the OMP, and separated by three well-conserved stretches (residues 37-130, 211-289, and 488-530 of fig. 3 alignment) probably associated with important biological functions.

In spite of the extreme genetic variability, the folding pattern of the envelope glycoprotein is probably constant. Indeed the position of virtually all of the cysteine residues is conserved within the different isolates (fig. 3 and 5), and the only three variable cysteines fall either in the signal peptide or in the very C-terminal part of the TMP. The hypervariable domains of the OMP are bounded by conserved cysteines, suggesting that they may represent loops attached to the common folding pattern. Also the calculated hydropathic profiles (Kyte and Doolittle, 1982) of the different envelope proteins are remarkably conserved (not shown).

About half of the potential N-glycosylation sites, Asn-X-Ser/Thr, found in the envelopes of the Zairian isolates map to the same positions in LAV_{BRU} (17/26 for LAV_{ELI} and 17/28 for LAV_{MAL}). The other sites appear to fall within variable domains of env,

suggesting the existence of differences in the extent of envelope glycosylation between different isolates.

Other viral proteins : Of the three other identified viral proteins, the p27 encoded by orf F, 3' of env (Allan et al., 1985b) is the most variable (fig. 4). The proteins encoded by orfs Q and S of the central region are remarkable by their absence of insertions/deletions. Surprisingly, a high frequency of aminoacids substitutions, comparable to that observed in env, is found for the product of orf S (trans-activating factor). On the other hand, the protein encoded by orf Q is no more variable than gag. Also noticeable is the lower variation of the proteins encoded by the central regions of LAV_{ELI} and LAV_{MAL}.

DISCUSSION

With the availability of the complete nucleotide sequence from five independant isolates, some general features of the AIDS virus genetic variability are now emerging. Firstly, its principal cause are point mutations very often resulting in amino-acid substitutions, and which are more frequent in the 3' part of the genome (orf S, env and orf F). Like all RNA viruses, the retroviruses are thought to be highly subject to mutations caused by errors of the RNA polymerases during their replication, since there is no proofreading, of this step (Holland et al., 1982 ; Steinhauer and Holland, 1986).

Another source of genetic diversity are insertions/deletions. From the figure 3 alignments, insertional events seem to be implicated in most of the cases, since otherwise deletions should have occurred in independant isolates at the precisely the same location. Furthermore, upon analyzing these insertions, we have observed that they most often represent one of the two copies of a direct repeat (fig. 6). Some are perfectly

conserved like the 36 bp repeat in the gag-pol overlap of LAV_{BRU} (fig. 6-a) ; others carry point mutations resulting in aminoacid substitutions, and as a consequence, they are more difficult to observe, though clearly present, in the hypervariable domains of env (cf. fig. 6-g and -h). As noted for point mutations, env gene and orf F also appear as more susceptible to that form of genetic ariation than the rest of the genome. The degree of conservation of these repeats must be related to their date of occurrence in the analyzed sequences : the more degenerated, the more ancient. A very recent divergence of LAV_{BRU} and HTLV3 is suggested by with extremely low number of mismatched AA between their homologous proteins. However, one of the LAV_{BRU} repeats (located in the Hyl domain of env, fig. 6-f) is not present in HTLV3, indicating that this generation of tandem repeats is a rapid source of genetic diversity. We have found no traces of such a phenomenon, even when comparing very closely related viruses, such as the Mason-Pfizer monkey virus, MPMV (Sonigo et al., 1986), and an immunosuppressive simian virus, SRV-1 (Power et al., 1986). Insertion or deletion of one copy of a direct repeat have been occasionally reported in mutant retroviruses (Shimotohno and Temin, 1981 ; Darlix, 1986), but the extent at which we observe this phenomenon is unprecedented.

The molecular basis of these duplications is unclear, but could be the "copy-choice" phenomenon, resulting from the diploidy of the retroviral genome (Varmus and Swanstrom, 1984 ; Clark and Mak, 1983). During the synthesis of the first-strand of the viral DNA, jumps are known to occur from one RNA molecule to another, especially when a break or a stable secondary structure is present on the template ; an inaccurate re-initiation on the other RNA template could

result in the generation (or the elimination) of a short direct repeat.

Genetic variability, and subsequent antigenic modifications, have often been developed by micro-organisms as a means to escape the host's immune response, either by modifying their epitopes during the course of the infection, as in trypanosomes (Borst and Cross, 1982), or by generating a large repertoire of antigens, as observed in influenza virus (Webster et al., 1982). As the human AIDS virus is related to animal lentiviruses (Sonigo et al., 1985 ; Chiu et al., 1985), its genetic variability could be a source of antigenic variation, as can be observed during the course of the infection by the ovine lentivirus visna (Scott et al., 1979 ; Clements et al., 1980) or by the equine infectious anemia virus (EIAV, Montelaro et al., 1984). However, a major discrepancy with these animal models is the extremely low, if any, neutralizing activity of the sera of individuals infected by the AIDS virus, whether they are healthy carriers, displaying minor symptoms or afflicted with AIDS (Weiss et al., 1985 ; Clavel, et al., 1985). Furthermore, even for the visna virus the exact role of antigenic variation in the pathogenesis is unclear (Thormar et al., 1983 ; Lutley et al., 1983). We rather feel that genetic variation represents a general selective advantage for lentiviruses by allowing an adaptation to different environments, for example by modifying their tissue or host tropisms. In the particular case of the AIDS virus, rapid genetic variations are tolerated, especially in the envelope ; they could allow the virus to get adapted to different "micro-environments" of the membrane of their principal target cells, namely the T4 lymphocytes. These "micro-environments" could result from the immediate vicinity of the virus receptor to polymorphic surface proteins, differring

either between individuals or betwwen clones of lymphocytes.

Conserved domains in the AIDS virus envelope.

Since the proteins of most of the isolates are antigenically cross-reactive, the genotypic differences do not seem to affect the sensitivity of actual diagnostic tests, based upon the detection of antibodies to the AIDS virus and using purified virions as antigens. They nevertheless have to be considered for the development of the "second-generation" tests, that are expected to be more specific, and will use smaller synthetic or genetically-engineered viral antigens. The identification of conserved domains in the highly immunogenic envelope glycoprotein, and also the core structural proteins (gag), is very important for these tests. The conserved stretch found at the end of the OMP and the beginning of the TMP (490-620, fig. 3) could be a good candidate, since a bacterial fusion protein containing this domain was well-detected by AIDS patients sera (Chang et al., 1985).

The envelope, specifically the OMP, mediates the interaction between a retrovirus and its specific cellular receptor (DeLarco and Todaro, 1976 ; Robinson et al., 1980). In the case of the AIDS virus, in vitro binding assays have shown the interaction of the envelope glycoprotein gp110 with the T4 cellular surface antigen (McDougal et al., 1986), already thought to be, or closely associated to, the virus receptor (Klatzmann et al., 1984 ; Dagleish et al., 1984). Identification of the AIDS virus envelope domains that are responsible for this interaction (receptor-binding domains) appears as fundamental for understanding of the host-viral interactions, but also for designing a protective vaccine, since an immune response against these epitopes could possibly elicit neutralizing antibodies. As the

AIDS virus receptor is at least partly formed of a constant structure, the T4 antigen, the binding site of the envelope is unlikely to be exclusively encoded by domains undergoing drastic genetic changes between isolates, even if these could be implicated in some kind of an "adaptation". One, or several of the conserved domains of the OMP (residues 37-130, 211-289, and 488-530 of fig. 3 alignment) brought together by the folding of the protein, must play a part in the virus-receptor interaction, and this can be explored with synthetic or genetically-engineered peptides derived from these domains, either by direct binding assays, or indirectly by assaying the neutralizing activity of specific antibodies raised against them.

African AIDS viruses

Zaire and the neighbouring countries of Central Africa are considered as an area of endemic for the AIDS virus infection, and the possibility that the virus has emerged in Africa has became a subject of intense controversy (see Norman, 1985). From the present study, it is clear that the genetic organization of Zairian isolates is the same as that of american isolates, thereby indicating a common origin. The very important sequence differences observed between the proteins are consistent with a divergent evolutionary process. In addition, the two African isolates are mutually more divergent than the American isolates already analyzed ; as far as that observation can be extrapolated, it suggests a longer evolution of the virus in Africa, and is also consistent with the fact that a larger fraction of the population is exposed than in developed countries.

A novel human retrovirus with morphology and biologocal properties (cytopathogenicity, T4 tropism) similar to those of LAV, but nevertheless clearly

genetically and antigenically distinct from that latter, was recently isolated from two patients with AIDS originating from Guinea Bissau, West-Africa (Clavel et al., 1986). In the neighbouring Senegal the population seems exposed to a retrovirus also distinct from LAV, but apparently non pathogenic (Barin et al., 1985 ; Kanki et al., 1986). Both of these novel African retroviruses seem to be antigenically related to the simian T-cell lymphotropic virus, STLV-III, shown to be widely present in healthy African green monkeys and other simian species (Kanki et al. 1985). This raises the possibility of a large group of African primate lentiviruses, ranging from the apparently non-pathogenic simian viruses to the LAV-type viruses. Their precise relationship will only be known after their complete genetic characterization, but it is already very likely that they have evolved from a common progenitor. The important genetic variability we have observed between isolates of the AIDS virus in Central Africa is probably a hallmark of this entire group, and may account for the apparently important genetic divergence between its members (loss of cross-antigenicity in the envelopes). In this sense the conservation of the tropism for the T4 lymphocytes suggests that it is a major advantage aquired by these retroviruses.

EXPERIMENTAL PROCEDURES

Virus isolations

LAV_{ELI} and LAV_{MAL} were isolated from the peripheral blood lymphocytes of the patients as described (Barré-Sinoussi et al., 1983) ; briefly, the lymphocytes were fractionated and co-cultivated with phytohaemagglutinin-stimulated normal human lymphocytes in the presence of interleukin 2 and anti-alpha interferon serum. Viral production was assessed by cell-free reverse transcriptase (RT) activity assay in the

cultures and by electron microscopy.

Molecular cloning

Normal donor lymphocytes were acutely infected (10^4 cpm of RT activity/10^6 cells) as described (Barré-Sinoussi et al., 1983), and total DNA was extracted at the beginning of the RT activity peak. For LAV_{ELI}, a lambda library using the L47-1 vector (Loenen and Brammar, 1982) was constructed by partial HindIII digestion of the DNA as already described (Alizon et al., 1984). For LAV_{MAL}, DNA from infected cells was digested to completion with HindIII and the 9-10kb fraction was selected on 0.8 % low melting point agarose gel and ligated into L47-1 HindIII arms. About 5.10^5 plaques for LAV_{ELI} and 2.10^5 for LAV_{MAL}, obtained by <u>in vitro</u> packaging (Amersham) were plated on <u>E. coli</u> LA101 and screened <u>in situ</u> under stringent conditions,using the 9 kb SacI insert of the clone lambda J19 (Alizon et al., 1984) carrying most of the LAV_{BRU} genome as probe. Clones displaying positive signals were plaque-purified and propagated on <u>E. coli</u> C600 recBC, and two recombinant phages carrying the complete genetic information of LAV_{ELI} (E-H12) and LAV_{MAL} (M-H11) were further characterized by restriction mapping.

Nucleotide sequence strategy

Viral fragments derived from E-H12 and M-H11 were sequenced by the dideoxy chain terminator procedure (Sanger et al., 1977) after "shotgun" cloning in the M13mp8 vector (Messing and Viera, 1982), as previously described (Sonigo et al., 1985). The viral genome of LAV_{ELI} is 9176 nucleotides, that of LAV_{MAL} 9229 nucleotides long. Each nucleotide was determined from more than 5 independent clones on average. Complete nucleotide sequences are not presented in this article for obvious reasons of space limitation but are freely available upon request to the authors, until they are released

through sequence data banks.

LEGEND OF THE FIGURES

Figure 1 : Restriction map analysis of AIDS virus isolates:

A/ Restriction maps of the inserts of phage lambda clones derived from cells infected with LAV_{ELI} (E-H12) and LAV_{MAL} (M-H11). The schematic genetic organization of the AIDS virus has been drawn above the maps. The LTRs are indicated by solid boxes. A:Aval-B:Bam HI-Bg:BgIII-E:EcoRI - H:HindIII - Hc:HincII - K:KpnI-N:Nde I-P:PstI-S:SacI-X:XbaI. Asterisks indicate the HindIII cloning sites in lambda L47-1 vector.

B/Comparison of the sites for seven restriction enzymes in six isolates : the prototype AIDS virus LAV_{BRU}, LAV_{MAL} and LAV_{ELI} ; Z1, Z2, Z3 are Zairian isolates with published restriction maps (Benn et al., 1985). Restriction sites are represented by the following symbols : BglII ; EcoRI ; HincII ; HindIII ; KpnI ; NdeI ; SacI.

Figure 2 : Conservation of the genetic organization of the central region in AIDS virus isolates.

Stop codons in each phse are represented as vertical bars. Vertical arrows indicate possible AUG initiation codons. Splice acceptor (A) and donor (D) sites identified in subgenomic viral mRNA (Muesing et al., 1985) are shown below the graphic of LAV_{BRU}, and corresponding sites in LAV_{ELI} and LAW_{MAL} are indicated. PPT indicates the repeat of the polypurine tract flanking the 3'LTR. As observed in LAV_{BRU} (Wain-Hobson et al., 1985), the PPT is repeated 256 nucleotides 5' to the end of the pol gene in both our sequences, but this repeat is degenerated at two positions in LAV_{ELI}.

Figure 3 : Alignment of the protein sequences of four AIDS virus isolates.

Isolate LAV_{BRU} (Wain-Hobson et al., 1985) is taken as reference ; only differences with LAV_{BRU} are noted for ARV2 (Sanchez-Pescador et al., 1985) and the two Zairian isolates LAV_{MAL} and LAV_{ELI}. A minimal number of gaps (-) was introduced in the alignments. The NH_2-termini of $p25^{gag}$ and $p18^{gag}$ are indicated (Sanchez-Pescador, 1985). The potential cleavage sites in the envelope precursor (Allan et al., 1985a ; diMarzo-Veronese, 1985) separating the signal peptide (SP), the outer membrane protein (OMP) and the transmembrane protein (TMP) are indicated as vertical arrows ; conserved cysteines are indicated by black circles and variable cysteines are boxed. The one letter code for amino acids is : A:Ala ; C:Cys ; D:Asp ; E:Glu ; F:Phe ; G:Gly ; H :His ; I:Ile ; K:Lys ; L:Leu ; M:Met ; N:Asn ; P:Pro ; Q:Gln ; R:Arg ; S:Ser ; T:Thr ; V:Val; W:Trp ; Y:Tyr.

Figure 4 : Quantitation of the sequence divergence between homologous proteins of different isolates.

Part A of each table gives results deduced from two-by-two alignments using the proteins of LAV_{BRU} as reference, part B those of LAV_{ELI} as reference . Sources: Muesing et al., 1985 for HTLV-3 ; Sanchez-Pescador et al., 1985 for ARV 2 and Wain-Hobson et al., 1985 for LAV_{BRU}. For each case of the tables, the size in amino-acids of the protein (calculated from the first methionine residue, or from the beginning of the orf for pol) is given at the upper left part. Below are given the number of deletions (left) and insertions (right) necessary for the alignment. The large numbers in bold face represent the percentage of amino-acids substitutions (insertions/deletions being excluded). Two by two alignments were done with computer assistance Wilburg and Lipman, 1983), using a gag penalty of 1, K-tuple of

1, and window of 20, except for the hypervariable domains of env, where the number of gaps was made minimum, and which are essentially aligned as shown in fig. 3. The sequence of the predicted protein encoded by orf R of HTLV-3 has not been compared because of a premature termination relative to all other isolates.

Figure 5 : Variability of the AIDS virus envelope protein.

For each position x of the alignment of env (Fig. 3), variability V(x) was calculated as

$$V(x) = \frac{\text{number of different amino-acids at position } x}{\text{frequency of the most abundant amino-acid at position } x.}$$

Gaps in the alignments are considered as another amino-acid. For an alignment of 4 proteins, V(x) ranges from 1 (identical AA in the 4 sequences) to 16 (4 different AA). This type of representation has previously been used in a compilation of the AA sequence of immunoglobulins variable regions (Wu and Kabat, 1970). Vertical arrows indicate the cleavage sites ; asterisks represent potential N-glysosylation sites (N-X-S/T) conserved in all four isolates ; black triangles represent conserved cysteine residues. Black lozanges mark the three major hydrophobic domains. OMP : outermembrane protein ; TMP : transmembrane protein ; signal : signal peptide ; Hyl, 2, 3, : hypervariable domains.

Figure 6 : Direct repeats in the proteins of different AIDS virus isolates.

These examples are derived from the aligned sequences of gag (a, b), F (c,d) an env (e, f, g, h) shown in figure 3. The two elements of the direct repeat are boxed, while degenerated positions are underlined.

The invention thus pertains more specifically to the proteins, polypeptides or glycoproteins including the polypeptidic strucutres shown in the drawings. The

first and last amino-acid residues of these proteins, polypeptides or glycoproteins carry numbers computed from a first aminoacid of the open-reading frames concerned, although these numbers do not correspond exactly to those of the LAV_{ELI} or LAV_{MAL} proteins concerned, rather to those of the LAV_{BRU} corresponding proteins or sequences shown in figs. 3A, 3B and 3C. Thus a number corresponding to a "first amino-acid residue" of a LAV_{ELI} protein corresponds to the number of the first amino-acyl residue of the corresponding LAV_{BRU} protein which, in any of figs. 3A, 3B or 3C is in direct alignment with the corresponding first amino-acid of the LAV_{ELI} protein. Thus the sequences concerned can be read from figs. 7A-7J and 8A-8I, to the extent where they do not appear with sufficient clarity from Figs. 3A-3F.

The preferred protein sequences of this invention extend from the corresponding "first" and "last" amino-acid residues (reference is also made to the protein(s)- or glycoprotein(s)-portions including part of the sequences which follow :

OMP or gp110 proteins, including precursors :
1 to 530

OMP or gp110 without precursor :
34-530

Sequence carrying the TMP or gp41 protein :
531-877, particularly
680-700

well conserved stretches of OMP :
37-130,
211-289 and
488-530

well conserved stretch found at the end of the OMP and the beginning of TMP :
490-620.

Proteins containing or consisting of the "well conserved stretches"are of particular interest for the production of immunogenic compositions and (preferably in relation to the stretches of the <u>env</u> protein) of vaccine compositions against the LAV-viruses of class 1 as above-defined.

The invention concerns more particularly all the DNA fragments which have been more specifically referred to in the drawings and which correspond to open reading frames. It will be understood that the man skilled in the art will be able to obtain them all, for instance by cleaving an entire DNA corresponding to the complete genome of either LAV_{ELI} or of LAV_{MAL}, such as by cleavage by a partial or complet digestion thereof with a suitable restriction enzyme and by the subsequent recovery of the relevant fragments. The different DNAs disclosed above can be resorted to also as a source of suitable fragments. The techniques disclosed in PCT application for the isolation of the fragments which can then be included in suitable plasmids are applicable here too.

Of course other methods can be used. Some of them have been examplified in European Application Nr. 178,978 filed on September 17, 1985. Reference is for instance made to the following methods.

a) DNA can be transfected into mammalian cells with appropriate selection markers by a variety of techniques, calcium phosphate precipitation, polyethylene glycol, protoplast-fusion, etc..

b) DNA fragments corresponding to genes can be cloned into expression vectors for E. coli, yeast- or mammalian cells and the resultant proteins purified.

c) The provival DNA can be "shot-gunned" (fragmented) into procaryotic expression vectors to generate fusion polypeptides. Recombinant producing

antigenically competent fusion proteins can be identified by simply screening the recombinants with antibodies against LAV antigens.

The invention further refers more specifically to DNA recombinants, particularly modified vectors, including any of the preceding DNA sequences and adapted to transform corresponding microorganisms or cells, particularly eucaryotic cells such as yeasts, for instance saccharomyces cerevisiae, or higher eucaryotic cells, particularly cells of mammals, and to permit expression of said DNA sequences in the corresponding microorganisms or cells. General methods of that type have been recalled in the abovesaid PCT international patent aplication PCT/EP 85/00548 filed on October 18, 1985.

More particularly the invention relates to such modified DNA recombinant vectors modified by the abovesaid DNA sequences and which are capable of transforming higher eucaryotic cells particularly mammalian cells. Preferably any of the abovesaid sequences are placed under the direct control of a promoter contained in said vectors and which is recognized by the polymerases of said cells, such that the first nucleotide codons expressed correspond to the first triplets of the above-defined DNA-sequences. Accordingly this invention also relates to the corresponding DNA fragments which can be obtained from genomas of LAV_{ELI} or LAV_{MAL} or corresponding cDNAs by any appropriate method. For instance such a method comprises cleaving said LAV genomas or cDNAs by restriction enzymes preferably at the level of restriction sites surrounding said fragments and close to the opposite extremities respectively thereof, recovering and identifying the fragments sought according to sizes, if need be checking their restriction maps or nucleotide sequences (or by reaction with

monoclonal antibodies specifically directed against epitopes carried by the polypeptides encoded by said DNA fragments), and further if need be, trimming the extremities of the fragment, for instance by an exonucleolytic enzyme such as Bal31, for the purpose of controlling the desired nucleotide-sequences of the extremities of said DNA fragments or, conversely, repairing said extremities with Klenow enzyme and possibly ligating the latter to synthetic polynucleotide fragments designed to permit the reconstitution of the nucleotide extremities of said fragments. Those fragments may then be inserted in any of said vectors for causing the expression of the corresponding polypeptide by the cell transformed therewith. The corresponding polypeptide can then be recovered from the transformed cells, if need be after lysis thereof, and purified, by methods such as electrophoresis. Needless to say that all conventional methods for performing these operations can be resorted to.

The invention also relates more specifically to cloned probes which can be made starting from any DNA fragment according to this invention, thus to recombinant DNAs containing such fragments, particularly any plasmids amplifiable in procaryotic or eucaryotic cells and carrying said fragments.

Using the cloned DNA fragments as a molecular hybridization probe - either by labelling with radionucleotides or with fluorescent reagents - LAV virion RNA may be detected directly in the blood, body fluids and blood products (e.g. of the antihemophylic factors such as Factor VIII concentrates) and vaccines, i.e. hepatitis B vaccine It has already been shown that whole virus can be detected in culture supernatants of LAV producing cells. A suitable method for achieving that detection comprises immobilizing virus onto a support,

e.g. nitrocellulose filters, etc., disrupting the virion and hybridizing with labelled (radiolabelled or "cold" fluorescent- or enzyme-labelled) probes. Such an approach has already been developed for Hepatitis B virus in peripheral blood (according to SCOTTO J. et al. Hepatology (1983), 3, 379-384).

Probes according to the invention can also be used for rapid screening of genomic DNA derived from the tissue of patients with LAV related symptoms, to see if the proviral DNA or RNA present in host tissue and other tissues can be related to that of LAV_{ELI} or LAV_{MAL}.

A method which can be used for such screening comprises the following steps : extraction of DNA from tissue, restriction enzyme cleavage of said DNA, electrophoresis of the fragments and Southern blotting of genomic DNA from tissues, subsequent hybridization with labelled cloned LAV proviral DNA. Hybridization in situ can also be used.

Lymphatic fluids and tissues and other non-lymphatic tissues of humans, primates and other mammalian species can also be screened to see if other evolutionnary related retrovirus exist. The methods referred to hereabove can be used, although hybridization and washings would be done under non stringent conditions.

The DNAs or DNA fragments according to the invention can be used also for achieving the expression of viral antigens of LAV_{ELI} or LAV_{MAL} for diagnostic purposes.

The invention relates generally to the polypeptides themselves, whether synthetized chemically isolated from viral preparation or expressed by the different DNAs of the inventions, particularly by the ORFs or fragments thereof, in appropriate hosts, particularly procaryotic or eucaryotic hosts, after

transformation thereof with a suitable vector previously modified by the corresponding DNAs.

More generally, the invention also relates to any of the polypeptide fragments (or molecules, particularly glycoproteins having the same polypeptidic backbone as the polypeptides mentioned hereabove) bearing an epitope characteristic of a protein or glycoprotein of LAV_{ELI} or LAV_{MAL}, which polypeptide or molecule then has N-terminal and C-terminal extremities respectively either free or, independently from each other, covalently bond to aminoacids other than those which are normally associated with them in the larger polypeptides or glycoproteins of the LAV virus, which last mentioned aminoacids are then free or belong to another polypeptidic sequence. Particularly the invention relates to hybrid polypeptides containing any of the epitope-bearing-polypeptides which have been defined more specifically hereabove, recombined with other polypeptides fragments normally foreign to the LAV proteins, having sizes sufficient to provide for an increased immunogenicity of the epitope-bearing-polypeptide yet, said foreign polypeptide fragments either being immunogenically inert or not interfering with the immunogenic properties of the epitope-bearing-polypeptide.

Such hybrid polypeptides which may contain from 5 up to 150, even 250 aminoacids usually consist of the expression products of a vector which contained ab initio a nucleic acid sequence expressible under the control of a suitable promoter or replicon in a suitable host, which nucleic acid sequence had however beforehand been modified by insertion therein of a DNA sequence encoding said epitope-bearing-polypeptide.

Said epitope-bearing-polypeptides, particularly those whose N-terminal and C-terminal aminoacids are

free, are also accessible by chemical synthesis, according to technics well known in the chemistry of proteins.

The synthesis of peptides in homogeneous solution and in solid phase is well known.

In this respect, recourse may be had to the method of synthesis in homogeneous solution described by Houbenweyl in the work entitled "Methoden der Organischen Chemie" (Methods of Organic Chemistry) edited by E. WUNSCH., vol. 15-I and II, THIEME, Stuttgart 1974.

This method of synthesis consists of successively condensing either the successive aminoacids in twos, in the appropriate order or successive peptide fragments previously available or formed and containing already several aminoacyl residues in the appropriate order respectively. Except for the carboxyl and amino-groups which will be engaged in the formation of the peptide bonds, care must be taken to protect beforehand all other reactive groups borne by these aminoacyl groups or fragments.However, prior to the formation of the peptide bonds, the carboxyl groups are advantageously activated, according to methods well known in the synthesis of peptides. Alternatively, recourse may be had to coupling reactions bringing into play conventional coupling reagents, for instance of the carbodiimide type, such as 1-ethyl-3-(3-dimethyl-aminopropyl)-carbodiimide. When the aminoacid group used carries an additional amine group (e.g. lysine) or another acid function (e.g. glutamic acid), these groups may be protected by carbobenzoxy or t-butyloxycarbonyl groups, as regards the amine groups, or by t-butylester groups, as regards the carboxylic groups. S imilar procedures are available for the protection of other reactive groups. for example, SH group (e.g. in cysteine) can be protected by an acetamidomethyl or paramethoxybenzyl group.

In the case of progressive synthesis, aminoacid by aminoacid, the synthesis starts preferably by the condensation of the C-terminal aminoacid with the aminoacid which corresponds to the neighboring aminoacyl group in the desired sequence and so on, step by step, up to the N-terminal aminoacid. Another preferred technique can be relied upon is that described by R.D. Merrifield in "solid phase peptide synthesis" (J. Am. Chem. Soc., 45, 2149-2154).

In accordance with the Merrifield process, the first C-terminal aminoacid of the chain is fixed to a suitable porous polymeric resin, by means of its carboxylic group, the amino group of said aminoacid then being protected, for example by a t-butyloxycarbonyl group.

When the first C-terminal aminoacid is thus fixed to the resin, the protective group of the amine group is removed by washing the resin with an acid, i.e. trifluoroacetic acid, when the protective group of the amine group is a t-butyloxycarbonyl group.

Then the carboxylic group of the second aminoacid which is to provide the second aminoacyl group of the desired peptidic sequence, is coupled to the deprotected amine group of the C-terminal aminoacid fixed to the resin. Preferably, the carboxyl group of this second aminoacid has been activated, for example by dicyclohexyl-carbodiimide, while its amine group has been protected, for example by a t-butyloxycarbonyl group. The first part of the desired peptide chain, which comprising the first two aminoacids, is thus obtained. As previously, the amine group is then deprotected, and one can further proceed with the fixing of the next aminoacyl group and so forth until the whole peptide sought is obtained.

The protective groups of the different side

groups, if any, of the peptide chain so formed can then be removed. The peptide sought can then be detached from the resin, for example, by means of hydrofluoric acid, and finally recovered in pure form from the acid solution according to conventional procedures.

As regards the peptide sequences of smallest size and bearing an epitope or immunogenic determinant, and more particularly those which are readily accessible by chemical synthesis, it may be required, in order to increase their in vivo immunogenic character, to couple or "conjugate" them covalently to a physiologically acceptable and non toxic carrier molecule.

By way of examples of carrier molecules or macromolecular supports which can be used for making the conjugates according to the invention, will be mentioned natural proteins, such as tetanic toxoid, ovalbumin, serum-albumins, hemocyanins, etc.. Synthetic macromolecular carriers, for example polysines or poly(D-L-alanine)-poly(L-lysine)s, can be used too.

Other types of macromolecular carriers which can be used, which generally have molecular weights higher than 20,000, are known from the literature.

The conjugates can be synthesized by known processes, such as described by Frantz and Robertson in "Infect. and Immunity", 33, 193-198 (1981), or by P.E. Kauffman in "Applied and Environmental Microbiology", October 1981 Vol. 42, n° 4, 611-614.

For instance the following coupling agents can be used : glutaric aldehyde, ethyl chloroformate, water-soluble carbodiimides (N-ethyl-N'(3-dimethylamino-propyl) carbodiimide, HCl), diisocyanates, bis-diazobenzidine, di- and trichloro-s-triazines, cyanogen bromides, benzaquinone, as well as coupling agents mentioned in "Scand. J. Immunol., 1978, vol. 8, p. 7-23 (Avrameas, Ternynck, Guesdon).

Any coupling process can be used for bonding one or several reactive groups of the peptide, on the one hand, and one or several reactive groups of the carrier, on the other hand. Again coupling is advantageously achieved between carboxyl and amine groups carried by the peptide and the carrier or vice-versa in the presence of a coupling agent of the type used in protein synthesis, i.e. 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide, N-hydroxybenzotriazole, etc.. Coupling between amine groups respectively borne by the peptide and the carrier can also be made with glutaraldehyde, for instance, according to the method described by BOQUET, P. et al. (1982) Molec. Immunol., 19, 1441-1549, when the carrier is hemocyanin.

The immunogenicity of epitope-bearing-peptides can also be reinforced, by oligomerisation thereof, for example in the presence of glutaraldehyde or any other suitable coupling agent. In particular, the invention relates to the water soluble immunogenic oligomers thus obtained, comprising particularly from 2 to 10 monomer units.

The glycoproteins, proteins and polypeptides (generally designated hereafter as "antigens" of this invention, whether obtained (by methods such as disclosed in the earlier patent applications referred to above) in a purified state from LAV_{ELI} or LAV_{MAL} virus preparations or - as concerns more particularly the peptides - by chemical synthesis, are useful in processes for the detection of the presence of anti-LAV antibodies in biological media, particularly biological fluids such as sera from man or animal, particularly with a view of possibly diagnosing LAS or AIDS.

Particularly the invention relates to an *in vitro* process of diagnosis making use of an envelope glycoprotein (or of a polypeptide bearing an epitope of

this glycoprotein of LAV_{ELI} or LAV_{MAL} for the detection of anti-LAV antibodies in the serums of persons who carry them. Other polypeptides - particular those carrying an epitope of a core protein - can be used too.

A preferred embodiment of the process of the invention comprises :

- depositing a predetermined amount of one or several of said antigens in the cups of a titration microplate ;
- introducing of increasing dilutions of the biological fluid, i.e. serum to be diagnosed into these cups ;
- incubating the microplate ;
- washing carefully the microplate with an appropriate buffer ;
- adding into the cups specific labelled antibodies directed against blood immunoglobulins and
- detecting the antigen-antibody-complex formed, which is then indicative of the presence of LAV antibodies in the biological fluid.

Advantageously the labelling of the anti-immunoglobulin antibodies is achieved by an enzyme selected from among those which are capable of hydrolysing a substrate, which substrate undergoes a modification of its radiation-absorption, at least within a predetermined band of wavelenghts. The detection of the substrate, preferably comparatively with respect to a control, then provides a measurement of the potential risks or of the effective presence of the disease.

Thus preferred methods immuno-enzymatic or also immunofluorescent detections, in particular according to the ELISA technique. Titrations may be determinations by immunofluorescence or direct or indirect immuno-enzymatic determinations. Quantitative titrations of antibodies on the serums studied can be made.

The invention also relates to the diagnostic

kits themselves for the in vitro detection of antibodies against the LAV virus, which kits comprise any of the polypeptides identified herein, and all the biological and chemical reagents, as well as equipment, necessary for peforming diagnostic assays. Preferred kits comprise all reagents required for carrying out ELISA assays. Thus preferred kits will include, in addition to any of said polypeptides, suitable buffers and anti-human immunoglobulins, which anti-human immunoglobulins are labelled either by an immunofluorescent molecule or by an enzyme. In the last instance preferred kits then also comprise a substrate hydrolysable by the enzyme and providing a signal, particularly modified absorption of a radiation, at least in a determined wavelength, which signal is then indicative of the presence of antibody in the biological fluid to be assayed with said kit.

It can of course be of advantage to use several proteins or polypeptides not only of both LAV_{ELI} and LAV_{MAL}, but also of any or both of them together with homologous proteins or polypeptides of earlier described viruses, e.g. of LAV_{BRU} or $HTLV_{III}$ or ARV, etc..

The invention also relates to vaccine compositions whose active principle is to be constituted by any of the antigen, i.e. the hereabove disclosed polypeptides whole antigens, of either LAV_{ELI} or LAV_{MAL}, or both, particularly the purified gp110 or immunogenic fragments thereof, fusion polypeptides or oligopeptides in association with a suitable pharmaceutical or physiologically acceptable carrier.

A first type of preferred active principle is the gp110 immunogen of said immunogens.

Other preferred active principles to be considered in that fields consist of the peptides containing less than 250 aminoacid units, preferably less than 150, particularly from 5 to 150 aminocid residues,

as deducible for the complete genomas of LAV_{ELI} and LAV_{MAL} and even more preferably those peptides which contain one or more groups selected from Asn-X-Ser and Asn-X-Ser as defined above. Preferred peptides for use in the production of vaccinating principles are peptides (a) to (f) as defined above. By way of example having no limitative character, there may be mentioned that suitable dosages of the vaccine compositions are those which are effective to elicit antibodies <u>in vivo</u>, in the host, particularly a human host. Suitable doses range from 10 to 500 micrograms of polypeptide, protein or glycoprotein per kg, for instance 50 to 100 micrograms per kg.

The different peptides according to this invention can also be used themselves for the production of antibodies, preferably monoclonal antibodies specific of the different peptides respectively. For the production of hybridomas secreting said monoclonal antibodies, conventional production and screening methods are used. These monoclonal antibodies, which themselves are part of the invention then provide very useful tools for the identification and even determination of relative proportions of the different polypeptides or proteins in biological samples, particularly human samples containing LAV or related viruses.

The invention further relates to the hosts (procaryotic or eucaryotic cells) which are transformed by the above mentioned recombinants and which are capable of expressing said DNA fragments.

Finally the invention also concerns vectors for the transformation fo eucaryotic cells of human origin, particularly lymphocytes, the polymerase of which are capable of recognizing the LTRs of LAV. Particularly said vectors are characterized by the presence of a LAV LTR therein, said LTR being then

active as a promoter enabling the efficient transcription and translation in a suitable host of a DNA insert coding for a determined protein placed under its controls.

Needless to say that the invention extends to all variants of genomes and corresponding DNA fragments (ORFs) having substantially equivalent properties, all of said genomes belonging to retroviruses which can be considered as equivalents of LAV.

It must be understood that the claims which follow are also intended to cover all equivalents of the products (glycoproteins, polypeptides, DNAs, etc..) whereby an equivalent is a product, i.e. a polypeptide which may distinguish from a determined one defined in any of said claims, say through one or several aminoacids, while still having substantially the same immunological or immunogenic properties. A similar rule of equivalency shall apply to the DNAs, it being understood that the rule of equivalency will then be tied to the rule of equivalency pertaining to the polypeptides which they encode.

It will also be understood that all the litterature referred to hereinbefore or hereinafter, and all patent applications or patents not specifically identified herein but which form counterparts of those specifically designated herein must be considered as incorporated herein by reference.

It should further be mentioned that the invention further relates to immunogenic compositions containing preferably not only any of the polypepdides more specifically identified above and which have the aminoacid-sequences of LAV_{ELI} and LAV_{MAL} which have been identified, but corresponding peptidic sequences to previously defined LAV proteins too.

In that respect the invention relates more

particularly to the particular polypeptides which have the sequences corresponding more specifically to the LAV_{BRU} sequences which have been referred to earlier, i.e. the sequences extending between the following first and last aminoacids, of the LAV_{BRU} proteins themselves, i.e. the polypeptides having sequences contained in the LAV_{BRU} OMP or LAV_{BRU} TMP or sequences extending over both, particularly those extending from between the following positions of the aminoacids included in the <u>env</u> open reading frame of the LAV_{BRU} genome,

1-530
34-530

and more preferably

531-877, particularly
680-700
37-130
211-289
488-530
490-620.

These different sequences can be used for any of the above defined purposes and in any of the compositions which have been disclosed.

Finally the invention also relates to the different antibodies which can be formed specifically against the different peptides which have been disclosed herein, particularly to the monoclonal antibodies which recognize them specifically. The corresponding hybridomas which can be formed starting from spleen cells previously immunized with such peptides which are fused with appropriate myeloma cells and selected according to standard procedures also form part of the invention.

Phage λ clone E-H12 derived from LAV_{ELI} infected cells has been deposited at the "Collection Nationale des Cultures de Micro-organismes" (National Collection of Cultures of Microorganisms) (CNCM) of the

Pasteur Institute of Paris France, under n° I-550 on May 9th, 1986.

Phage λ clone M-H11 derived from LAV_{MAL} infected cells has been deposited at the CNCM under n° I-551 on May 9th, 1986.

REFERENCES

Alizon, M., Sonigo, P., Barré-Sinoussi, F., Chermann, J.C., Tiollais, P., Montagnier, L. & Wain-Hobson, S. (1984). Molecular cloning of lymphadenopathy-associated virus. Nature 312, 757-760.

Allan, J.S., Coligan, J.E., Barin, F., McLane, M.F., Sodroski, J.G., Rosen, C.A., Haseltine, W.A., Lee, T.H., & Essex, M. (1985a). Major glycoprotein antigens that induce antibodies in AIDS patients. Science 228, 1091-1094.

Allan, J.S., Coligan, J.E., Lee, T.H., McLane, M.F., Kanki, P.J., Groopman, J.E., & Essex, M. (1985b). A new HTLV-III/LAV antigen detected by antibodies from AIDS patients. Science 230, 810-813.

Arya, S.K., Guo, C., Josephs, S.F., & Wong-Staal, F. (1985). Trans-activator gene of human T-lymphotropic virus type III (HTLV-III). Science 229, 69-73.

Bailey, A.C., Downing, R.G., Cheinsong-Popov, R., Tedder, R.C., Dalgleish, A.G., & Weiss, R.A.(1985). HTLV-III serology distinguishes atypical and endemic Kaposi's sarcoma in Africa. Lancet I, 359-361.

Barin, F., M'Boup, S., Denis, F., Kanki, P., Allan, J.S., Lee, T.M., & Essex, M. (1985). Serological evidence for virus related to simian T lymphotropic retrovirus in residents of West Africa. Lancet II, 1387-1389.

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Brun-Vézinet, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk of acquired immune deficiency syndrome (AIDS). Science 220, 868-870.

Benn, S., Rutledge, R., Folks, T., Gold, J., Baker, L., McCormick, J., Feorino, P., Piot, P., Quinn T. & Martin, M. (1985). Genomic heterogeneity of AIDS retroviral isolates from North America and Zaire. Science 230, 949-951.

Borst, P., & Cross, G.A.M. (1982). Molecular basis for trypanosome antigenic variation. Cell 29, 291-303.

Brun-Vézinet, F., Rouzioux, C., Montagnier, L., Chamaret, S., Gruest, J., Barré-Sinoussi, F., Geroldi, D., Chermann, J.C., McCormick, J., Mitchell, S., Piot, P., Taelmann, H., Minlangu, K.B., Wobin, O., Mbendi, N., Mazebo, P., Kalambayi, K., Bridts, C., Desmyter, J., Feinsod, F., & Quinn T.C. (1984). Prevalence of antibodies to lymphadenopathy-associated virus in African patients with AIDS. Science 226, 453-456.

Chang, N.T., Chanda, P.K., Barone, A.D., McKinney, S., Rhodes, D.P., Tam, S.H. Shearman, C.W., Huang, J. & Chang, T.W. (1985). Expression in Escherichia coli of open reading frame gene segments of HTLV-III. Science 228, 93-96.

Chiu, I.M., Yaniv, A., Dahlberg, J.E., Gazit, A., Skuntz, S.F., Tronick, S.R. & Aaronson, S. A. (1985). Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. Nature 317, 366-368.

Clark, S.P., & Mak, T.W. (1984). Fluidity of a retrovirus genome. J.Virol. 50, 759-765.

Clavel, F., Klatzmann, D., & Montagnier, L. (1985). Deficient neutralizing capacity of sera from patients with AIDS or related syndromes. Lancet I, 879-880.

Clavel, F., Brun-Vézinet, F., Guetard, D., Chamaret, S., Laurent, A., Rouzioux, C., Rey, M., Katlama, C., Rey, F., Champelinaud, J.L., Nina, J.S., Mansinho, K., Santos-Ferreira, M.O., Klatzmann, D., & Montagnier, L.(1986). LAV type II : a second retrovirus associated with AIDS in West-Africa. C.R.Acad.Sci.Paris 302, 485-488.

Clements, J.E., Narayan, O., Griffin, D.E. and Johnson, R.T. (1980). Genomic changes associated with antigenic variation of visna virus during persistent infection. Proc. Natl. Acad. Sci. USA 77, 4454-4458.

Clumeck, N., Sonnet, J., Taelman, M., Mascart-Lemone, F., De Bruyère, M., Van de Perre, P., Dasnoy, J., Marcelis, L., Lamy, M., Jonas, C., Eyckmans, L.,Noel, H., Vanhaeverbeek, M. & Butzler, J. P. (1984). Acquired immune deficiency syndrome in African patients. N. Engl. J. Med., 310, 492-497.

Dalgleish, A.G., Beverley, P.C., Clapham, P.R., Crawford, D.H., Greaves, M.F. & Weiss, R.A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 312, 763-767.

Darlix, J.L. (1986) Control of Rous sarcoma virus RNA translation and packaging by the 5' and 3' untranslated sequences. J.Mol.Biol., in the press.

DeLarco, J. & Todaro, G.J. (1976). Membrane receptors of murine leukemia viruses : characterization using the purified viral envelope glycoprotein, gp71. Cell 8, 365-371.

DiMarzoVeronese, F., DeVico, A.L., Copeland, T.D., Oroszlan, S., Gallo, R.C., & Sarngadharan, M.G. (1985). Characterization of gp 41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. Science 229, 1403-1405.

Ellrodt, A., Barré-Sinoussi, F., Le Bras, P., Nugeyre, M.T., Brun-Vézinet, F., Rouzioux, C., Segond, P., Caquet, R., Montagnier, L. & Chermann, J.C. (1984). Isolation of human T-lymphotropic retrovirus (LAV) from Zairan married couple, one with AIDS, one with prodromes. Lancet I, 1383-1385.

Hahn, B.H., Shaw, G.M., Arya, S.U., Popovic, M., Gallo, R.C. & Wong-Staal, F. (1984). Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature 312, 166-169.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., & Van de Pol, S. (1982). Rapid evolution of RNA genomes. Science 215, 1577-1585.

Kan, N.C., Franchini, G., Wong-Staal, F., Dubois, G.C., Robey, W.G., Lautenberger, J.A. & Papas, T.S. (1986). Identification of HTLV-III/LAV sor gene product and detection of antibodies in human sera. Science 231, 1553-1555.

Kanki, P.J., Alroy, J. & Essex, M. (1985). Isolation of T-lymphotropic retroviruses from wild-caught African Green Monkeys. Science 230, 951-954.

Kanki, P.J., Barin, F., M'Boup, S., Allan, J.S., Romet-Lemonne, J.L., Markink, R., McLane, M.F., Lee, T.H., Arbeille, B., Denis, F. & Essex, M. (1986). New human T-lymphotropic retrovirus related to simian T-lymphotropic virus type III ($STLV-III_{AGM}$). Science, 232, 238-243.

Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.C. & Montagnier, L. (1984). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312, 767-768.

Kyte, J. & Doolittle, R. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

Koch, W., Hunsmann, G. & Friedrich, R. (1983). Nucleotide sequence of the envelope gene of Friend murine leukemia virus. J. Virol., 45, 1-9.

Lee, T.H., Coligan, J.E., Allan, J.S., McLane, M.F., Groopman, J.E. & Essex, M. (1986). A new HTLV III/LAV protein encoded by a gene found in cytopathic retroviruses. Science 231, 1546-1549.

Loenen, W.A.M. & Brammar, W.J. (1980). A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene 10, 249-259.

Lutley, R., Petursson, G., Palsson, P.A., Georgsson, G., Klein, J., & Nathanson, N. (1983). Antigenic drift in visna : virus variation during long-term infection of icelandic sheep. J. Gen. Virol. 64, 1433-1440.

MacDougal, J.S., Kennedy, M.S., Sligh, J.M., Cort, S.P., Mawle, A. & Nicholson, J.K.A. (1986). Binding of HTLV-III/LAV to T4+ cells by a complex of the 110 k viral protein and the T4 molecule. Science 231, 382-385.

Messing, J. and Viera, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene 19, 269-276.

Montagnier, L. (1985). Lymphadenopathy-associated virus : from molecular biology to pathogenicity. Ann.Intern.Med. 103, 689-693.

Montagnier, L., Clavel, F., Krust, B., Chamaret, S., Rey, F., Barré-Sinoussi, F. & Chermann, J.C. (1985). Identification and antigenicity of the major envelope glycoprotein of lymphadenopathy-associated virus. Virology 144, 283-289.

Montelaro, R.C., Parekh, B., Orrego, A. & Issel, C.J. (1984). Antigenic variation during persistent infection by equine infectious anemia virus, a retrovirus. J. Biol. Chem., 259, 10539-10544.

Muesing, M.A., Smith, D.M., Cabradilla, C.D., Benton, C.V., Lasky, L.A. & Capon, D.J. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retroviruses. Nature 313, 450-458.

Norman, C. (1985). Politics and science clash on African AIDS. Science 230, 1140-1142.

Piot, P., Quinn, T.C., Taelmann, H., Feinsod, F.M., Minlangu, K.B., Wobin, O., Mbendi, N., Mazebo, P., Ndongi, K., Stevens, W., Kalambayi, K., Mitchell, S., Bridts, C. & McCormick, J.B. (1984). Acquired immunodeficiency syndrome in

Power, M.D., Marx, P.A., Bryant, M.L., Gardner, M.B., Barr, P.J. & Luciw, P.A. (1986). Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science 231, 1567-1572.

Rabson, A.B. & Martin, M.A. (1985). Molecular organization of the AIDS retrovirus. Cell 40, 477-480.

Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, A., Whitchorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Pearson, M.L., Lautenbergen, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. & Wong-Staal, F. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 313, 277-284.

Robinson, P.J.G., Hunsmann, G., Schneider, J. & Schirrmacher, V. (1980). Possible cell surface receptor for Friend murine leukemia virus is isolated with viral envelop glycoprotein complexes. J. Virol., 36, 291-294.

Rosen, C.A., Sodroski, J.G. & Haseltine, W.A. (1985). The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. Cell 41, 813-823.

Sanchez-Pescador, R. Power, M.D., Barr, P.J., Steimer, K.S., Stemfeien, M.M., Brown-Shimer, S.L., Gee, W.W., Bernard, A., Randolph, A., Levy, J.A., Dina, D. & Luciw, P.A., (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science 227, 484-492.

Sanger, F., Nicklen, S. & Coulsen, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Scott, J.V., Stowring, L., Haase, A.T., Narayan, O. & Vigne, R. (1979). Antigenic variation in visna virus. Cell 18, 321-327.

Shimotohno, K., & Temin, H. (1982). Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. J.Virol. 41, 163-171.

Sodroski, J., Patarca, R., Rosen C., Wong-Staal, F. & Haseltine, W. (1985). Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. Science 229, 74-77.

Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. & Wain-Hobson, S. (1985). Nucleotide sequence of the visna lentivirus : Relationship to the AIDS virus. Cell 42, 369-382.

Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson S. (1986). Nucleotide sequence of Mason-Pfizer Monkey virus : an immunosuppressive D-type retrovirus. Cell, in the press.

Steinhauer, D.A., & Holland, J.H. (1986). Direct method for quantitation of extreme polymerase error frequencies at selected single base in viral RNA. J.Virol. 57, 219-228.

Thormar, H., Barshatsky, M.R., Arnesen, K., & Kozlowski, P.B. (1983). The emergence of antigenic variants is a rare event in long-term visna virus infection <u>in vivo</u>. J. Gen. Virol. 64, 1427-1432.

Van de Perre, P., Rouvroy, D., Lepage, P., Bogaerts, J., Kestelyn, P., Kayihigi, J., Hekker, A.C., Butzler, J.P. & Clumeck, N. (1984). Acquired immunodeficiency syndrome in Rwanda. Lancet II, 62-65.

Varmus, H. & Swanstrom, R. (1984). Replication of retroviruses. In Molecular biology of the tumor viruses : RNA tumor viruses. R. Weiss, N. Teich, H. Varmus, J. Coffin, eds. (Cold Spring Harbor Laboratory, New York), vol. I, pp. 369-512.

Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. Cell 40, 9-17.

Webster, R.G., Laver, W.G., Air, G.M. & Schild, G.C. (1982). Molecular mechanisms of variation in influenza viruses. Nature 296, 115-121.

Weiss, R.A. (1984). Human T-cell retroviruses. In Molecular biology of the tumor viruses : RNA tumor viruses. R. Weiss, N. Teich, H. Varmus, J. Coffin, eds. (Cold Spring Harbor Laboratory, New York), vol II : supplement, pp. 405-485.

Weiss, R.A., Clapham, P.R., Cheinsong-Popov, R., Dalgleish, A.G., Carne, C.A., Weller, I.A.D. & Tedder, R.C. (1985). Neutralization of human T-lymphotropic virus type III by sera of AIDS and AIDS-risk patients. Nature, 316, 69-72.

Wilburg, W.J., & Lipman, D.J. (1983). Rapid similarity searches of nucleic acid and protein data banks. Proc.Natl.Acad.Sci.USA 80, 726-730.

Wu, T.T., & Kabat, E.A. (1970). An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implications for antibody complementarity. J.Exp.Med. 132, 211-250.

We claim :

1. The virus LAV_{ELI} whose RNA corresponds to the cDNA of figs. 7A-7J.

2. The virus LAV_{MAL} whose RNA corresponds to the cDNA of figs. 8A-8I.

3. The cDNA of figs. 7A-7J or parts thereof.

4. The cDNA of figs. 8A-8I or parts thereof.

5. DNA recombinants containing at least part of the cDNA of claim 3 or 4.

6. A probe containing a cloned nucleic acid according to any of claims 3 to 5.

7. The method for identifying the presence in a host tissue of a virus or provirus related to either LAV_{ELI} or LAV_{MAL} which comprises hybridizing DNA obtained from said tissue with a probe according to claim 6 and detecting the presence of said virus or provirus in said tissue according as hybridization with said probe is detected or not.

8. A peptide, protein, or parts thereof encoded by open reading frames of the DNA sequences of claim 3 or 4 or fragments thereof.

9. A peptide of claim 8 which corresponds to any of the stretches extending respectively
from aminoacyl residue 37 to aminoacyl residue 130,
or from aminoacyl residue 211 to aminoacyl residue 289,
or from aminoacyl residue 488 to aminoacyl residue 530,
of fig. 3.

10. A peptide of claim 8 which corresponds to the stretch extending from the aminoacyl residue 490 to the aminocyl residue 620 of fig. 3

11. A portion of a protein or glycoprotein whose aminoacid sequence includes all or part of the sequences which follow :

OMP or gp110 proteins, including precursors :
1 to 530
OMP or gp110 without precursor :
34-530
Sequence carrying the TMP or gp41 protein :
531-877, particularly
680-700

well conserved stretches of OMP :
37-130,
211-289 and
488-530

well conserved stretch found at the end of the OMP and the beginning of TMP :
490-620.

12. A method for the <u>in vitro</u> detection of the presence of antibodies directed against LAV_{ELI} or LAV_{MAL} or against related viruses in human body fluids which comprises contacting said body fluids with antigens obtained from the viruses of claims 1 or 2 or consisting of peptides according to any of claims 8 to 10 and detecting the immunological reaction between said antigens and said antibodies.

13. The method claim 11 which comprises :

- depositing a predetermined amount of one or several of said antigens in the cups of a titration microplate ;
- introducing of increasing dilutions of the biological fluid, i.e. serum to be diagnosed into these cups ;
- incubating the microplate ;
- washing carefully the microplate with an appropriate buffer ;
- adding into the cups specific labelled antibodies directed against blood immunoglobulins and
- detecting the antigen-antibody-complex formed, which is then indicative of the presence of LAV antibodies in the biological fluid.

14. A diagnostic kit for the *in vitro* detection of antibodies against the viruses of claims 1 or 2 or viruses related therewith, which contain an antigen obtained from said viruses or consisting of a peptide according to any of claims 8 to 11, and the biological and chemical reagents, as well as equipment, necessary for performing diagnostic assays.

15. An immunogenic composition containing an antigen of the viruses of claim 1 or 2, or both or of any immunogenic peptide encoded by the RNAs of said viruses or by part thereof in association with a suitable pharmaceutically or physiologically acceptable carrier.

16. The immunogenic composition of claim 15 wherein said peptide is the gp110 enveloppe glycoprotein or part thereof.

17. The immunogenic composition of claim 16 which contains a protein or glycoprotein whose aminoacid sequence includes all or part of any of the sequences which follow :

OMP or gp110 proteins, including precursors :
1 to 530

OMP or gp110 without precursor :
34-530

Sequence carrying the TMP or gp41 protein :
531-877, particularly
680-700

well conserved stretches of OMP :
37-130,
211-289 and
488-530

well conserved stretch found at the end of the OMP and the beginning of TMP :
490-620.

18. The antibodies, particularly monoclonal antibodies, formed against any of the peptides, proteins or glycoproteins of any of claims 8 to 11.

19. The cells transformed with a DNA recombinant according to claim 5.

LAV.ELI

R
GGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTGGCTAGCTAGGGAACCCAC

R ←→ U5
TGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGT
100

GTGACTCTGGTAACTAGAGATCCCTCAGACCCCTTTAGTCAGAGTGGAAAATCTCTAGCA

U5←
GTGGCGCCCGAACAGGGACCTGAAAGCGAAAGTAGAACCAGAGGAGCTCTCTCGACGCAG
200

GACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCAGCGACTGGTGAGTACGCT
300

→GAG
MetGlyAlaArgAlaSerValLeuSer
AAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAA

GlyGlyLysLeuAspLysTrpGluLysIleArgLeuArgProGlyGlyLysLysLysTyr
GCGGGGGAAAATTAGATAAATGGGAAAAAATTCGGTTACGGCCAGGAGGAAAGAAAAAAT
400

ArgLeuLysHisIleValTrpAlaSerArgGluLeuGluArgTyrAlaLeuAsnProGly
ATAGACTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATATGCACTTAATCCTG

LeuLeuGluThrSerGluGlyCysLysGlnIleIleGlyGlnLeuGlnProAlaIleGln
GCCTTTTAGAAACATCAGAAGGCTGTAAACAAATAATAGGGCAGCTACAACCAGCTATTC
500

ThrGlyThrGluGluLeuArgSerLeuTyrAsnThrValAlaThrLeuTyrCysValHis
AGACAGGAACAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTAC
600

LysGlyIleAspValLysAspThrLysGluAlaLeuGluLysMetGluGluGluGlnAsn
ATAAAGGAATAGATGTAAAAGACACCAAGGAAGCTTTAGAAAAGATGGAGGAAGAGCAAA

LysSerLysLysLysAlaGlnGlnAlaAlaAlaAspThrGlyAsnAsnSerGlnValSer
ACAAAAGTAAGAAAAAAGGCACAGCAAGCAGCAGCTGACACAGGAAACAACAGCCAGGTCA
700

GlnAsnTyrProIleValGlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerPro
GCCAAAATTATCCTATAGTGCAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCAC

ArgThrLeuAsnAlaTrpValLysValIleGluGluLysAlaPheSerProGluValIle
CTAGAACTTTGAACGCATGGGTAAAAGTAATAGAAGAAAAGGCTTTCAGCCCAGAAGTAA
800

ProMetPheSerAlaLeuSerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsn
TACCCATGTTTTCAGCATTATCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAA
900

ThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAla
ACACAGTGGGGGGACATCAAGCAGCCATGCAAATGCTAAAAGAGACCATCAATGAAGAAG

AlaGluTrpAspArgLeuHisProValHisAlaGlyProIleAlaProGlyGlnMetArg
CTGCAGAATGGGATAGGTTACATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGA
1000

GluProArgGlySerAspIleAlaGlyThrThrSerThrLeuGlnGluGlnIleAlaTrp
GAGAACCAAGGGGAAGTGATATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGCAT

MetThrSerAsnProProIleProValGlyGluIleTyrLysArgTrpIleIleValGly
GGATGACAAGTAACCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATTGTGG
1100

LeuAsnLysIleValArgMetTyrSerProValSerIleLeuAspIleArgGlnGlyPro
GATTAAATAAAATAGTAAGAATGTATAGCCCTGTCAGCATTTTGGACATAAGACAGGGAC
1200

LysGluProPheArgAspTyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAla
CAAAGGAACCTTTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAG
. . . . . .

SerGlnAspValLysAsnTrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAsp
CTTCACAGGATGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAG
. . . 1300 . .

CysLysThrIleLeuLysAlaLeuGlyProGlnAlaThrLeuGluGluMetMetThrAla
ATTGCAAGACTATCTTAAAAGCATTGGGACCACAGGCTACACTAGAAGAAATGATGACAG
. . . . . .

CysGlnGlyValGlyGlyProSerHisLysAlaArgValLeuAlaGluAlaMetSerGln
CATGTCAGGGAGTGGGGGGCCCAGCCATAAAGCAAGAGTTCTGGCTGAGGCAATGAGCC
. 1400 . . . .

AlaThrAsnSerValThrThrAlaMetMetGlnArgGlyAsnPheLysGlyProArgLys
AAGCAACAAATTCAGTTACTACAGCAATGATGCAGAGAGGCAATTTTAAGGGCCCAAGAA
. 1500

IleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaLysAsnCysArgAlaPro
AAATTATTAAGTGTTTCAATTGTGGCAAAGAAGGGCACATAGCAAAAAATTGCAGGGCCC
. . . . . .

ArgLysLysGlyCysTrpArgCysGlyLysGluGlyHisGlnLeuLysAspCysThrGlu
CTAGGAAAAAGGGCTGTTGGAGATGTGGAAAGGAAGGACACCAACTAAAAGATTGCACTG
. → POL . . 1600 . .

PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaGlyGluLeu
ArgGlnAlaAsnPheLeuGlyArgIleTrpProSerHisLysGlyArgProGlyAsnPhe
AGAGACAGGCTAATTTTTTAGGGAGAATTTGGCCTTCCCACAAGGGAAGGCCGGGAACT
. . . . . .

SerProLysGlnThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArg
LeuGlnSerArgProGluProThrAlaProProAlaGluSerPheGlyPheGlyGluGlu
TTCTCCAAAGCAGACCAGAGCCAACAGCCCCACCAGCAGAGAGCTTCGGGTTTGGGGAAG
. 1700 . . . .

AspAsnProLeuSerLysThrGlyAlaGluArgGlnGlyThrValSerPheAsnPhePro
IleThrProSerGlnLysGlnGluGlnLysAspLysGluLeuTyrProLeuThrSerLeu
AGATAACCCCCTCTCAAAAACAGGAGCAGAAAGACAAGGAACTGTATCCTTTAACTTCCC
. . . GAG ← . . 1800

GlnIleThrLeuTrpGlnArgProLeuValAlaIleLysIleGlyGlyGlnLeuLysGlu
LysSerLeuPheGlyAsnAspProLeuSerGln
TCAAATCACTCTTTGGCAACGACCCCTTGTCGCAATAAAAATAGGGGGACAGCTAAAGGA
. . . . . .

AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLys
AGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAA
. . . 1900 . .

TrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGln
ATGGAAACCAAAAATGATAGGGGGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCA
. . . . . .

IleProIleGluIleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrPro
AATACCCATAGAAATCTGTGGACAGAAAGCTATAGGTACAGTATTAGTAGGACCTACGCC
. 2000 . . . .

ValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIle
TGTCAACATAATCGGAAGAAATTTGTTGACCCAGATTGGCTGCACTTTAAATTTTCCAAT
. 2100

SerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLys
TAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAA
. . . . . .

GlnTrpProLeuThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGlu
ACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAACAGAAATTTGTACAGATATGGA
. . . 2200 . .

```
LysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAla
AAAGGAAGGAAAAATTTCAAGAATTGGGCCTGAAAATCCATACAATACTCCAATATTTGC

IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys
CATAAAGAAAAAAGACAGTACCAAGTGGAGAAAATTAGTAGATTTCAGAGAACTTAATAA
                         2300
ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLys
GAGAACTCAAGATTTCTGGGAAGTTCAATTAGGAATACCGCATCCTGCAGGGCTGAAAAA
                                                        2400
LysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGlu
GAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCCTTAGATGA

AspPheArgLysTyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIle
AGATTTTAGGAAATATACCGCCTTTACCATATCTAGTATAAACAATGAGACACCAGGGAT
                                           2500
ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSer
TAGATATCAGTACAATGTGCTTCCACAGGGATGGAAAGGATCACCGGCAATATTCCAAAG

SerMetThrLysIleLeuGluProPheArgLysGlnAsnProGluMetValIleTyrGln
TAGCATGACAAAAATCTTAGAGCCCTTTAGAAAACAAAATCCAGAAATGGTTATCTATCA
                         2600
TyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIle
ATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGGACAAAAAT
                                                        2700
GluLysLeuArgGluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGln
AGAGAAATTAAGAGAACATCTATTGAGGTGGGGATTTACCAGACCAGATAAAAAACATCA

LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln
GAAAGAACCCCCATTTCTTTGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACA
                                           2800
SerIleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGlu
GTCTATAAAACTGCCAGAAAAGGAGAGCTGGACTGTCAATGATATACAGAACTTAGTGGA

ArgLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeu
GAGATTAAACTGGGCAAGCCAGATTTATCCAGGAATTAAAGTAAGACAATTATGTAAACT
                         2900
LeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGlu
CCTTAGGGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAATTAGA
                                                        3000
LeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSer
ACTGGCAGAAAACAGGGAAATTTTAAAAGAACCAGTACATGGAGTGTATTATGACCCATC

LysAspLeuIleAlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyr
AAAAGACTTAATAGCAGAAATACAGAAACAAGGGCACGGCCAATGGACATACCAAATTTA
                                           3100
GlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThr
TCAAGAACCATTTAAAAATCTGAAAACAGGAAAGTATGCAAGAATGAGGGGTGCCCACAC

AsnAspValLysGlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIle
TAATGATGTAAAGCAATTAGCAGAGGCAGTGCAAAGAATATCCACAGAAAGCATAGTGAT
                         3200
TrpGlyArgThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrp
ATGGGGAAGGACTCCTAAATTTAGACTACCCATACAAAAGGAAACATGGGAAACATGGTG
                                                        3300
```

```
 AlaGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu
GGCAGAGTATTGGCAAGCCACTTGGATTCCTGAGTGGGAATTTGTCAATACCCCTCCTTT
          .         .         .         .         .         .
 ValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrVal
AGTAAAATTATGGTACCAGTTAGAGAAGGAACCCATAATAGGAGCAGAAACTTTCTATGT
          .         .         .       3400        .         .
 AspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGly
AGATGGGGCAGCTAATAGAGAGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGG
          .         .         .         .         .         .
 ArgGlnLysValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIle
AAGACAGAAAGTTGTCCCTTTGACTGACACGACAAATCAGAAGACTGAGTTACAAGCAAT
          .       3500        .         .         .         .
 AsnLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAla
TAATCTAGCCTTGCAGGATTCGGGATTAGAAGTAAACATAGTAACAGATTCACAATATGC
          .         .         .         .         .       3600
 LeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIle
ATTAGGAATCATTCAAGCACAACCAGATAAGAGTGAATCAGAGTTAGTCAATCAAATAAT
          .         .         .         .         .         .
 GluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIle
AGAGCAGTTAATAAAAAAGGAAAAGGTTTACCTGGCATGGGTACCAGCACACAAAGGAAT
          .         .         .       3700        .         .
 GlyGlyAsnGluGlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeu
TGGAGGAAATGAACAAGTAGATAAATTAGTCAGTCAAGGAATCAGGAAAGTACTATTTTT
          .         .         .         .         .         .
 AspGlyIleAspLysAlaGlnGluGluHisGluLysTyrHisAsnAsnTrpArgAlaMet
GGATGGAATAGATAAGGCTCAAGAAGAACATGAGAAATATCACAACAATTGGAGAGCAAT
          .       3800        .         .         .         .
 AlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSerCysAspLys
GGCTAGTGATTTTAACCTACCACCCGTGGTAGCAAAAGAAATAGTAGCTAGCTGTGATAA
          .         .         .         .         .       3900
 CysGlnLeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGlyIleTrpGln
ATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCA
          .         .         .         .         .         .
 LeuAspCysThrHisLeuGluGlyLysValIleLeuValAlaValHisValAlaSerGly
ATTAGATTGTACACACTTAGAAGGAAAAGTTATCCTGGTAGCAGTTCATGTAGCCAGTGG
          .         .         .       4000        .         .
 TyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeu
CTATATAGAAGCAGAAGTTATTCCAGCAGAAACAGGGCAGGAAACAGCATATTTTCTTTT
          .         .         .         .         .         .
 LysLeuAlaGlyArgTrpProValLysValValHisThrAspAsnGlySerAsnPheThr
AAAATTAGCAGGAAGATGGCCAGTAAAAGTAGTACATACAGACAATGGCAGCAATTTCAC
          .       4100        .         .         .         .
 SerAlaAlaValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPheGlyIlePro
CAGTGCTGCAGTTAAGGCCGCCTGTTGGTGGGCAGGTATCAAACAGGAATTTGGAATTCC
          .         .         .         .         .       4200
 TyrAsnProGlnSerGlnGlyValValGluSerMetAsnLysGluLeuLysLysIleIle
CTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATTAAAGAAAATTAT
          .         .         .         .         .         .
 GlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAlaValPheIle
AGGACAGGTAAGAGATCAAGCTGAACATCTTAAGACAGCAGTACAAATGGCAGTATTCAT
          .         .         .       4300        .         .
 HisAsnPheLysArgArgArgGlyIleGlyGlyTyrSerAlaGlyGluArgIleIleAsp
CCACAATTTTAAAAGAAGAAGGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAATAGA
          .         .         .         .         .         .
```

IleIleAlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleIleLysIleGlnAsn
CATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTATAAAAATTCAAAA
4400

PheArgValTyrTyrArgAspSerArgAspProIleTrpLysGlyProAlaLysLeuLeu
TTTTCGGGTTTATTACAGAGACAGCAGAGATCCAATTTGGAAAGGACCAGCAAAGCTCCT
4500

TrpLysGlyGluGlyAlaValValIleGlnAspLysSerAspIleLysValValProArg
CTGGAAAGGTGAAGGGGCAGTAGTAATACAAGACAAGAGTGACATAAAGGTAGTACCAAG

ArgLysValLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAla
Q
MetGluAsnArgTrpGlnValMetIleValTrpGln
AAGAAAAGTAAAGATTATTAGGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGC
POL 4600

SerArgGlnAspGluAsp
ValAspArgMetArgIleLysThrTrpLysSerLeuValLysHisHisMetTyrValSer
AAGTAGACAGGATGAGGATTAAAACATGGAAAAGTTTAGTAAAACACCATATGTATGTTT

LysLysAlaAsnArgTrpPheTyrArgHisHisTyrGluSerProHisProLysIleSer
CAAAGAAAGCTAACAGATGGTTTTATAGACATCACTATGAAAGCCCCCACCCAAAAATAA
4700

SerGluValHisIleProLeuGlyGluAlaArgLeuValIleLysThrTyrTrpGlyLeu
GTTCAGAAGTACACATCCCACTAGGAGAAGCTAGACTGGTAATAAAAACATATTGGGGTC
4800

HisThrGlyGluArgGluTrpHisLeuGlyGlnGlyValSerIleGluTrpArgLysArg
TGCATACAGGAGAAAGAGAATGGCATCTGGGTCAGGGAGTCTCCATAGAATGGAGGAAAA

ArgTyrSerThrGlnValAspProGlyLeuAlaAspGlnLeuIleHisMetTyrTyrPhe
GGAGATATAGCACACAAGTAGACCCTGGCCTGGCAGACCAACTAATTCATATGTATTATT
4900

AspCysPheSerGluSerAlaIleArgLysAlaIleLeuGlyAspIleValSerProArg
TTGATTGTTTTTCAGAATCTGCTATAAGAAAAGCCATATTAGGAGATATAGTTAGTCCTA

CysGluTyrGlnAlaGlyHisAsnLysValGlySerLeuGlnTyrLeuAlaLeuThrAla
GGTGTGAGTATCAAGCAGGACATAACAAGGTAGGATCCCTACAGTATTTGGCACTAACAG
5000

LeuIleAlaProLysGlnIleLysProProLeuProSerValArgLysLeuThrGluAsp
CATTAATAGCACCAAAACAGATAAAGCCACCTTTGCCTAGTGTTAGGAAGCTAACAGAAG
R 5100

MetGluGlnAlaProAlaAspGlnGlyProGlnArgGluProTyrAsnGluTrpAla
ArgTrpAsnLysProGlnGlnThrArgGlyHisArgGlySerHisThrMetAsnGlyHis
ATAGATGGAACAAGCCCCAGCAGACCAGGGGCCACAGAGGGAGCCATACAATGAATGGGC

Q
LeuGluLeuLeuGluGluLeuLysSerGluAlaValArgHisPheProArgIleTrpLeu
ATTAGAGCTTTTAGAGGAGCTTAAGAGTGAAGCTGTTAGACATTTTCCTAGGATATGGCT
5200

HisSerLeuGlyGlnHisIleTyrGluThrTyrGlyAspThrTrpValGlyValGluAla
CCATAGCTTAGGACAACATATTTATGAAACTTATGGGGATACCTGGGTAGGAGTTGAAGC

IleIleArgIleLeuGlnGlnLeuLeuPheIleHisPheArgIleGlyCysGlnHisSer
TATAATAAGAATACTGCAACAATTACTGTTTATTCATTTCAGAATTGGGTGTCAACATAG
5300 S R

ArgIleGlyIleIleArgGlnArgArgAlaArgAsnGlySerSerArgSer
MetAspProValAspProAsnLeuGlu
CAGAATAGGCATTATTCGACAGAGAAGAGCAAGAAATGGATCCAGTAGATCCTAACCTAG
5400

```
 ProTrpAsnHisProGlySerGlnProArgThrProCysAsnLysCysHisCysLysLys
AGCCCTGGAACCATCCAGGAAGTCAGCCTAGGACTCCTTGTAACAAGTGTCATTGTAAAA
        .         .         .         .         .         .
 CysCysTyrHisCysProValCysPheLeuAsnLysGlyLeuGlyIleSerTyrGlyArg
AGTGTTGCTATCATTGCCCAGTTTGCTTCTTAAACAAAGGCTTAGGCATCTCCTATGGCA
        .         .         .       5500        .         .
 LysLysArgArgGlnArgArgGlyProProGlnGlyGlyGlnAlaHisGlnValProIle
GGAAGAAGCGGAGACAGCGACGAGGACCTCCTCAAGGCGGTCAGGCTCATCAAGTTCCTA
        S<-       .         .         .         .         .
 ProLysGln
TACCAAAGCAGTAAGTAGTACATGTAATGCAACCTTTAGGGATAATAGCAATAGCAGCAT
        .       5600        .         .         .         .
TAGTAGTAGCAATAATACTAGCAATAGTTGTGTGGACCATAGTATTCATAGAATATAGAA
        .         .         .         .         .       5700
GGATAAAAAAGCAAAGGAGAATAGACTGTTTACTTGATAGAATAACAGAAAGAGCAGAAG
        ->ENV     .         .         .         .         .
         MetArgAlaArgGlyIleGluArgAsnCysGlnAsnTrpTrpLysTrpGly
ACAGTGGCAATGAGAGCGAGGGGGATAGAGAGAAATTGTCAAAACTGGTGGAAATGGGGC
        .         .         .       5800        .         .
IleMetLeuLeuGlyIleLeuMetThrCysSerAlaAlaAspAsnLeuTrpValThrVal
ATCATGCTCCTTGGGATATTGATGACCTGTAGTGCTGCAGACAATCTGTGGGTCACAGTT
        .         .         .         .         .         .
TyrTyrGlyValProValTrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAla
TATTATGGGGTGCCTGTATGGAAGGAAGCAACCACCACTCTATTTTGTGCATCAGATGCT
        .       5900        .         .         .         .
LysSerTyrGluThrGluAlaHisAsnIleTrpAlaThrHisAlaCysValProThrAsp
AAATCATATGAAACAGAGGCACATAATATCTGGGCCACACATGCCTGTGTACCCACGGAC
        .         .         .         .         .       6000
ProAsnProGlnGluIleAlaLeuGluAsnValThrGluAsnPheAsnMetTrpLysAsn
CCCAACCCACAAGAAATAGCACTGGAAAATGTGACAGAAAACTTTAACATGTGGAAAAAT
        .         .         .         .         .         .
AsnMetValGluGlnMetHisGluAspIleIleSerLeuTrpAspGlnSerLeuLysPro
AACATGGTGGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAACCA
        .         .         .       6100        .         .
CysValLysLeuThrProLeuCysValThrLeuAsnCysSerAspGluLeuArgAsnAsn
TGTGTAAAATTAACCCCACTCTGTGTCACTTTAAACTGTAGTGATGAATTGAGGAACAAT
        .         .         .         .         .         .
GlyThrMetGlyAsnAsnValThrThrGluGluLysGlyMetLysAsnCysSerPheAsn
GGCACTATGGGGAACAATGTCACTACAGAGGAGAAAGGAATGAAAAACTGCTCTTTCAAT
        .       6200        .         .         .         .
ValThrThrValLeuLysAspLysLysGlnGlnValTyrAlaLeuPheTyrArgLeuAsp
GTAACCACAGTACTAAAAGATAAGAAGCAGCAAGTATATGCACTTTTTTATAGACTTGAT
        .         .         .         .         .       6300
IleValProIleAspAsnAspSerSerThrAsnSerThrAsnTyrArgLeuIleAsnCys
ATAGTACCAATAGACAATGATAGTAGTACCAATAGTACCAATTATAGGTTAATAAATTGT
        .         .         .         .         .         .
AsnThrSerAlaIleThrGlnAlaCysProLysValSerPheGluProIleProIleHis
AATACCTCAGCCATTACACAGGCTTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACAT
        .         .         .       6400        .         .
TyrCysAlaProAlaGlyPheAlaIleLeuLysCysArgAspLysLysPheAsnGlyThr
TATTGTGCCCCAGCTGGTTTTGCGATTCTAAAGTGTAGAGATAAGAAGTTCAATGGAACA
        .         .         .         .         .         .
GlyProCysThrAsnValSerThrValGlnCysThrHisGlyIleArgProValValSer
GGCCCATGCACAAATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTGGTGTCA
        .       6500        .         .         .         .
```

ThrGlnLeuLeuLeuAsnGlySerLeuAlaGluGluGluValIleIleArgSerGluAsn
ACTCAACTGCTGTTGAATGGCAGTCTAGCAGAAGAAGAGGTCATAATTAGATCCGAAAAT
6600
LeuThrAsnAsnAlaLysAsnIleIleAlaHisLeuAsnGluSerValLysIleThrCys
CTCACAAACAATGCTAAAAACATAATAGCACATCTTAATGAATCTGTAAAAATTACCTGT

AlaArgProTyrGlnAsnThrArgGlnArgThrProIleGlyLeuGlyGlnSerLeuTyr
GCAAGGCCCTATCAAAATACAAGACAAAGAACACCTATAGGACTAGGGCAATCACTCTAT
6700
ThrThrArgSerArgSerIleIleGlyGlnAlaHisCysAsnIleSerArgAlaGlnTrp
ACTACAAGATCAAGATCAATAATAGGACAAGCACATTGTAATATTAGTAGAGCACAATGG

SerLysThrLeuGlnGlnValAlaArgLysLeuGlyThrLeuLeuAsnLysThrIleIle
AGTAAAACTTTACAACAAGTAGCTAGAAAATTAGGAACCCTTCTTAACAAAACAATAATA
6800
LysPheLysProSerSerGlyGlyAspProGluIleThrThrHisSerPheAsnCysGly
AAGTTTAAACCATCCTCAGGAGGGGACCCAGAAATTACAACACACAGTTTTAATTGTGGA
6900
GlyGluPhePheTyrCysAsnThrSerGlyLeuPheAsnSerThrTrpAsnIleSerAla
GGGGAATTCTTCTACTGTAATACATCAGGACTGTTTAATAGTACATGGAATATTAGTGCA

TrpAsnAsnIleThrGluSerAsnAsnSerThrAsnThrAsnIleThrLeuGlnCysArg
TGGAATAATATTACAGAGTCAAATAATAGCACAAACACAAACATCACACTCCAATGCAGA
7000
IleLysGlnIleIleLysMetValAlaGlyArgLysAlaIleTyrAlaProProIleGlu
ATAAAACAAATTATAAAGATGGTGGCAGGCAGGAAAGCAATATATGCCCCTCCTATCGAA

ArgAsnIleLeuCysSerSerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyIle
AGAAACATTCTATGTTCATCAAATATTACAGGGCTACTATTGACAAGAGATGGTGGTATA
7100
AsnAsnSerThrAsnGluThrPheArgProGlyGlyGlyAspMetArgAspAsnTrpArg
AATAATAGTACTAACGAGACCTTTAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGA
7200
SerGluLeuTyrLysTyrLysValValGlnIleGluProLeuGlyValAlaProThrArg
AGTGAATTATATAAATATAAGGTAGTACAAATTGAACCACTAGGAGTAGCACCCACCAGG

AlaLysArgArgValValGluArgGluLysArgAlaIleGlyLeuGlyAlaMetPheLeu
GCAAAGAGAAGAGTGGTGGAAAGAGAAAAAAGAGCAATAGGATTAGGAGCTATGTTCCTT
7300
GlyPheLeuGlyAlaAlaGlySerThrMetGlyAlaArgSerValThrLeuThrValGln
GGGTTCTTGGGAGCAGCAGGAAGCACGATGGGCGCACGGTCAGTGACGCTGACGGTACAG

AlaArgGlnLeuMetSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGlu
GCCAGACAATTAATGTCTGGTATAGTGCAACAGCAAAACAATTTGCTGAGGGCTATAGAG
7400
AlaGlnGlnHisLeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgIle
GCGCAACAGCATCTGTTGCAACTCACGGTCTGGGGCATTAAACAGCTCCAGGCAAGAATC
7500
LeuAlaValGluArgTyrLeuLysAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGly
CTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTAGGAATTTGGGGTTGCTCTGGA

```
LysHisIleCysThrThrAsnValProTrpAsnSerSerTrpSerAsnArgSerLeuAsn
AAACACATTTGCACCACTAATGTGCCCTGGAACTCTAGTTGGAGTAATAGATCTCTAAAT
    .         .         .        7600         .         .
GluIleTrpGlnAsnMetThrTrpMetGluTrpGluArgGluIleAspAsnTyrThrGly
GAGATTTGGCAGAACATGACCTGGATGGAGTGGGAAAGAGAAATTGACAATTACACAGGC
    .         .         .         .         .         .
LeuIleTyrSerLeuIleGluGluSerGlnThrGlnGlnGluLysAsnGluLysGluLeu
TTAATATATAGCTTAATTGAGGAATCGCAGACCCAGCAAGAAAAGAATGAAAAAGAATTG
    .       7700        .         .         .         .
LeuGluLeuAspLysTrpAlaSerLeuTrpAsnTrpPheSerIleThrGlnTrpLeuTrp
TTGGAATTGGACAAGTGGGCAAGTTTGTGGAATTGGTTTAGCATAACACAATGGCTGTGG
    .         .         .         .         .        7800
TyrIleLysIlePheIleMetIleIleGlyGlyLeuIleGlyLeuArgIleValPheAla
TATATAAAAATATTCATAATGATAATAGGAGGCTTGATAGGTTTAAGAATAGTTTTTGCT
    .         .         .         .         .         .
ValLeuSerLeuValAsnArgValArgGlnGlyTyrSerProLeuSerPheGlnThrLeu
GTGCTTTCTTTAGTAAATAGAGTTAGGCAGGGATACTCACCTCTGTCGTTTCAGACCCTC
    .         .         .        7900         .         .
LeuProAlaProArgGlyProAspArgProGluGlyThrGluGluGluGlyGlyGluArg
CTCCCAGCCCCGAGGGGACCCGACAGGCCCGAAGGAACAGAAGAAGAAGGTGGAGAGCGA
    .         .         .         .         .         .
GlyArgAspArgSerValArgLeuLeuAsnGlyPheSerAlaLeuIleTrpAspAspLeu
GGCAGAGACAGATCCGTGAGATTGCTGAACGGATTCTCGGCACTTATCTGGGACGACCTG
    .       8000        .         .         .         .
ArgSerLeuCysLeuPheSerTyrHisArgLeuArgAspLeuIleLeuIleAlaValArg
CGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTAATCTTAATTGCAGTGAGG
    .         .         .         .         .        8100
IleValGluLeuLeuGlyArgArgGlyTrpAspIleLeuLysTyrLeuTrpAsnLeuLeu
ATTGTAGAACTTCTGGGACGCAGGGGGTGGGACATCCTCAAATATCTGTGGAATCTCCTA
    .         .         .         .         .         .
GlnTyrTrpSerGlnGluLeuArgAsnSerAlaSerSerLeuPheAspAlaIleAlaIle
CAGTATTGGAGTCAGGAACTGAGGAACAGTGCTAGTAGCTTGTTTGATGCCATAGCAATA
    .         .         .        8200         .         .
AlaValAlaGluGlyThrAspArgValIleGluIleIleGlnArgAlaCysArgAlaVal
GCAGTAGCTGAGGGGACAGATAGAGTTATAGAAATAATACAAAGAGCTTGCAGAGCTGTT
    .         .         .         .   W       .         .
LeuAsnIleProArgArgIleArgGlnGlyLeuGluArgSerLeuLeu         F
                                                      MetGlyGly
CTTAACATACCCAGAAGAATAAGACAGGGCTTAGAAAGGTCTTTACTTTAAAATGGGTGG
    .       8300        .         .         .         .
LysTrpSerLysSerSerIleValGlyTrpProAlaIleArgGluArgIleArgArgThr
CAAATGGTCAAAAAGTAGTATAGTGGGATGGCCTGCTATAAGGGAAAGAATAAGAAGAAC
    .         .         .         .         .        8400
AsnProAlaAlaAspGlyValGlyAlaValSerArgAspLeuGluLysHisGlyAlaIle
TAATCCAGCAGCAGATGGGGTAGGAGCAGTATCTCGAGACCTGGAAAAACATGGGGCAAT
    .         .         .         .         .         .
ThrSerSerAsnThrAlaSerThrAsnAlaAspCysAlaTrpLeuGluAlaGlnGluGlu
CACAAGTAGCAATACAGCAAGTACTAATGCTGACTGTGCCTGGCTAGAAGCACAAGAAGA
    .         .         .        8500         .         .
SerAspGluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLys
GAGCGACGAGGTGGGCTTTCCAGTCAGACCCCAGGTACCTTTAAGACCAATGACTTACAA
    .         .         .         .   U3    .         .
GluAlaLeuAspLeuSerHisPheLeuLysGluLysGlyGlyLeuGluGlyLeuIleTrp
AGAAGCTCTAGATCTCAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTTG
    .       8600        .         .         .         .
```

```
 SerLysLysArgGlnGluIleLeuAspLeuTrpValTyrAsnThrGlnGlyIlePhePro
GTCCAAAAAGAGACAAGAGATCCTTGATCTTTGGGTCTACAACACACAAGGCATCTTCCC
                                                        8700
 AspTrpGlnAsnTyrThrProGlyProGlyIleArgTyrProLeuThrPheGlyTrpCys
TGATTGGCAAAACTACACACCAGGGCCAGGGATCAGATATCCACTAACCTTTGGATGGTG

 TyrGluLeuValProValAspProGlnGluValGluGluAspThrGluGlyGluThrAsn
CTACGAGCTAGTACCAGTTGATCCACAGGAGGTAGAAGAAGACACTGAAGGAGAGACCAA
                                   8800
 SerLeuLeuHisProIleCysGlnHisGlyMetGluAspProGluArgGlnValLeuLys
CAGCTTGTTACACCCTATATGCCAGCATGGAATGGAGGACCCGGAGAGACAAGTGTTAAA

 TrpArgPheAsnSerArgLeuAlaPheGluHisLysAlaArgGluMetHisProGluPhe
ATGGAGATTTAACAGCAGACTAGCATTTGAGCACAAGGCCCGAGAGATGCATCCGGAGTT
               8900
 TyrLysAsn
CTACAAAAACTGATGACACCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGG
                                                        9000
GAGGCGTGGACTGGGCGGGACTGGGGAGTGGCTAACCCTCAGATGCTGCATATAAGCAGC
              U3    R
TGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATTTGAGCCTGGGAGCTCTCTG
                                   9100                B
GCTAGCTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAA
```

LAV.MAL

```
R
GGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGCAAGGAACCCACTG
                                 R  U5
CTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAAGCAGTGTGTGCCCATCTGTTGTGT
                                      100             U5
GACTCTGGTAACTAGAGATCCCTCAGACCACTCTAGACGGTGTAAAAATCTCTAGCAGTG

GCGCCCGAACAGGGACTTTAAAGTGAAAGTAACAGGGACTCGAAAGCGGAAGTTCCAGAG
            200
AAGTTCTCTCGACGCAGGACTCGGCTTGCTGAGGTGCACACAGCAAGAGGCGAGAGCGGC
                                                          300
                                                     GAG
                                                   MetGlyAlaArg
GACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAG

AlaSerValLeuSerGlyGlyLysLeuAspAlaTrpGluLysIleArgLeuArgProGly
AGCGTCAGTATTAAGCGGGGGAAAATTAGATGCATGGGAGAAAATTCGGTTAAGGCCAGG
                                      400
GlyLysLysLysTyrArgLeuLysHisLeuValTrpAlaSerArgGluLeuGluArgPhe
GGGAAAGAAAAAATATAGACTGAAACATTTAGTATGGGCAAGCAGGGAGCTGGAAAGATT

AlaLeuAsnProGlyLeuLeuGluThrGlyGluGlyCysGlnGlnIleMetGluGlnLeu
CGCACTTAACCCTGGCCTTTTAGAAACAGGAGAAGGATGTCAACAAATAATGGAACAGCT
            500
GlnSerThrLeuLysThrGlySerGluGluIleLysSerLeuTyrAsnThrValAlaThr
ACAATCAACTCTCAAGACAGGATCAGAAGAAATTAAATCATTATATAATACAGTAGCAAC
                                                          600
LeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAlaLeuAspLysIle
CCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACCAAGGAAGCGCTAGATAAAAT

GluGluIleGlnAsnLysSerArgGlnLysThrGlnGlnAlaAlaAlaAlaGlnGlnAla
AGAGGAAATACAAAATAAGAGCAGGCAAAAGACACAGCAGGCAGCAGCTGCACAGCAGGC
                                      700
AlaAlaAlaThrLysAsnSerSerSerValSerGlnAsnTyrProIleValGlnAsnAla
AGCAGCTGCCACAAAAAACAGCAGCAGTGTCAGTCAAAATTACCCCATAGTGCAAAATGC

GlnGlyGlnMetIleHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysVal
ACAAGGGCAAATGATACATCAGGCCATATCACCTAGGACTTTGAATGCATGGGTGAAAGT
            800
IleGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGly
AATAGAAGAAAAGGCTTTCAGCCCAGAAGTGATACCCATGTTCTCAGCATTATCAGAGGG
                                                          900
AlaThrProGlnAspLeuAsnMetMetLeuAsnIleValGlyGlyHisGlnAlaAlaMet
GGCCACCCCACAAGATTTAAATATGATGCTGAACATAGTTGGAGGACACCAGGCAGCTAT

GlnMetLeuLysAspThrIleAsnGluGluAlaAlaAspTrpAspArgValHisProVal
GCAAATGTTAAAAGATACCATCAATGAGGAAGCTGCAGACTGGGACAGGGTACATCCAGT
                                      1000
HisAlaGlyProIleProProGlyGlnMetArgGluProArgGlySerAspIleAlaGly
ACATGCAGGGCCTATTCCCCCAGGCCAGATGAGAGAACCAAGAGGAAGTGACATAGCAGG
```

Thr ~~ThrSerThr~~LeuGlnGluGlnIleGlyTrpMetThrSerAsnProProIleProVal
AACTACTAGTACCCTTCAAGAACAAATAGGATGGATGACAAGCAACCCACCTATCCCAGT
1100

GlyAspIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSer
GGGAGACATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAG
1200

ProValSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAsp
CCCTGTCAGCATTTTGGACATAAGACAAGGGCCAAAGGAACCTTTTAGAGACTATGTAGA

ArgPhePheLysThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThr
TAGGTTCTTTAAAACTCTCAGAGCTGAGCAAGCTACACAGGAGGTAAAAAATTGGATGAC
1300

GluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGly
AGAAACCTTGCTGGTCCAAAATGCGAATCCAGACTGTAAGACCATTTTAAAAGCATTAGG

ProGlyAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHis
ACCAGGGGCTACATTAGAAGAAATGATGACAGCATGCCAGGGAGTGGGAGGACCCAGTCA
1400

LysAlaArgValLeuAlaGluAlaMetSerGlnAlaThrAsnSerThrAlaAlaIleMet
TAAAGCAAGAGTTTTGGCTGAGGCAATGAGCCAAGCAACAAATTCAACTGCTGCCATAAT
1500

MetGlnArgGlyAsnPheLysGlyGlnLysArgIleLysCysPheAsnCysGlyLysGlu
GATGCAGAGAGGTAATTTTAAGGGCCAGAAAAGAATTAAGTGTTTCAACTGTGGCAAAGA

GlyHisLeuAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLys
AGGACACCTAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAA
1600 → POL
PhePheArgGluAsnLeu
GluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrp
GGAAGGACACCAAATGAAAGACTGCACTGAGAGACAGGCTAATTTTTTAGGGAAAATTTG

AlaPheProGlnGlyLysAlaArgGluPheProSerGluGlnThrArgAlaAsnSerPro
ProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPro-
GCCTTCCCACAAGGGAAGGCCAGGGAATTTCCTTCAGAGCAGACCAGAGCCAACAGCCCC
1700

ThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeuSerGluThrGlyAlaGlu
ProAlaGluSerPheGlyPheGlyGluGluIleLysProSerGlnLysGlnGluGlnLys
ACCAGCAGAGAGCTTCGGGTTTGGGGAGGAGATAAAACCCTCTCAGAAACAGGAGCAGAA
1800

ArgGlnGlyIleValSerPheSerPheProGlnIleThrLeuTrpGlnArgProValVal
AspLysGluLeuTyrProLeuAlaSerLeuLysSerLeuPheGlyAsnAspGlnLeuSer
AGACAAGGAATTGTATCCTTTAGCTTCCCTCAAATCACTCTTTGGCAACGACCAGTTGTC
GAG ←
ThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThr
Gln
ACAGTAAGAGTAGGAGGACAGCTAAAAGAAGCTCTATTAGACACAGGAGCAGATGATACA
1900

ValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGly
GTATTAGAAGAAATAAATTTGCCAGGAAAATGGAAACCAAAAATGATAGGGGGAATTGGA

GlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAla
GGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCT
2000

```
IleGlyThrIleLeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThr
ATAGGTACAATATTGGTAGGACCTACACCTGTCAACATAATTGGACGAAATATGTTGACT
         .         .         .         .         .      2100
GlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeu
CAGATTGGTTGTACTTTAAATTTTCCAATTAGTCCTATTGAGACTGTACCAGTAAAATTA
         .         .         .         .         .         .
LysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGluGluLysIleLys
AAGCCAGGGATGGATGGCCCAAGGGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAA
         .         .         .      2200         .         .
AlaLeuThrGluIleCysLysAspMetGluLysGluGlyLysIleLeuLysIleGlyPro
GCATTAACAGAAATTTGTAAAGATATGGAAAAGGAAGGAAAAATTTTAAAAATTGGGCCT
         .         .         .         .         .         .
GluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArg
GAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGCACTAAATGGAGA
         .      2300         .         .         .         .
LysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeu
AAATTAGTGAATTTCAGAGAGCTTAATAAAAGAACTCAAGATTTTTGGGAAGTTCAATTA
         .         .         .         .         .      2400
GlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGly
GGAATACCACATCCTGCTGGGTTGAAAAAGAAAAAATCAGTCACAGTATTGGATGTGGGG
         .         .         .         .         .         .
AspAlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIle
GATGCATATTTTTCAGTCCCTTTAGATGAAGATTTCAGGAAGTATACTGCATTCACTATA
         .         .         .      2500         .         .
ProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGly
CCCAGTATTAATAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTACCACAGGGA
         .         .         .         .         .         .
TrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArg
TGGAAAGGATCACCAGCAATATTCCAGAGTAGCATGACAAAAATCTTAGAACCCTTTAGA
         .      2600         .         .         .         .
ThrLysAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAsp
ACAAAAAATCCAGAAATAGTCATATACCAATACATGGATGATTTGTATGTAGGGTCTGAT
         .         .         .         .         .      2700
LeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGluHisLeuLeuLysTrp
TTAGAAATAGGACAACATAGAACAAAAATAGAGGAACTAAGAGAACATCTATTGAAATGG
         .         .         .         .         .         .
GlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyr
GGATTTACCACACCAGACAAAAAGCATCAGAAAGAACCCCCATTTCTTTGGATGGGGTAT
         .         .         .      2800         .         .
GluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysGluSerTrp
GAACTCCACCCTGACAAATGGACAGTGCAGCCTATACAACTGCCAGACAAGGAAAGCTGG
         .         .         .         .         .         .
ThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrPro
ACTGTCAATGATATACAGAAATTGGTGGGAAAACTAAATTGGGCAAGTCAGATTTATCCA
         .      2900         .         .         .         .
GlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIle
GGAATTAAAGTAAAGCAATTATGTAAACTCCTTAGGGGAGCAAAAGCACTAACAGACATA
         .         .         .         .         .      3000
ValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGlu
GTACCATTAACTGCAGAGGCAGAATTAGAATTGGCAGAGAACAGGGAAATTCTAAAAGAA
         .         .         .         .         .         .
```

```
ProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGln
CCAGTGCATGGGGTATATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAG
          .          .          .        3100       .         .
GlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyrLysAsnLeuLysThrGly
GGGCAAGGTCAATGGACATATCAAATATACCAAGAGCAATATAAAAATCTGAAAACAGGG
          .          .          .          .          .         .
LysTyrAlaArgIleLysSerAlaHisThrAsnAspValLysGlnLeuThrGluAlaVal
AAGTATGCAAGAATAAAGTCTGCCCACACTAATGATGTAAAACAATTAACAGAAGCAGTG
          .       3200          .          .          .         .
GlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArgLeuPro
CAAAAGATAGCCCAAGAAAGCATAGTAATATGGGGAAAAACTCCTAAATTTAGACTACCC
          .          .          .          .          .      3300
IleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGlnAlaThrTrpIlePro
ATACAAAAAGAAACATGGGAGGCATGGTGGACAGAATATTGGCAAGCCACCTGGATCCCT
          .          .          .          .          .         .
GluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGlu
GAATGGGAGTTTGTCAATACTCCTCCCCTAGTAAAACTATGGTACCAGTTAGAAACAGAA
          .          .          .        3400       .         .
ProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLys
CCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGAGAAACTAAAAAG
          .          .          .          .          .         .
GlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSerLeuThrGluThr
GGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAGGTTGTCTCCTTAACTGAAACA
          .       3500          .          .          .         .
ThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlySerGlu
ACAAATCAGAAGACTGAATTACAAGCAATCCACTTAGCTTTACAGGATTCAGGATCAGAA
          .          .          .          .          .      3600
ValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLys
GTAAACATAGTAACAGACTCACAGTATGCATTAGGGATTATTCAAGCACAACCAGATAAA
          .          .          .          .          .         .
SerGluSerGluIleValAsnGlnIleIleGluGlnLeuIleGlnLysAspLysValTyr
AGTGAATCAGAGATTGTTAATCAAATAATAGAGCAATTAATACAGAAGGACAAGGTCTAC
          .          .          .        3700       .         .
LeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuVal
CTGTCATGGGTACCAGCACACAAAGGGATTGGAGGAAATGAACAAGTAGATAAATTAGTC
          .          .          .          .          .         .
SerSerGlyIleArgLysValLeuPheLeuAspGlyIleAspLysAlaGlnGluGluHis
AGCAGTGGAATCAGAAAGGTACTATTTTTAGATGGGATAGATAAGGCTCAAGAAGAACAT
          .       3800          .          .          .         .
GluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeuProProIleVal
GAAAAATATCACAGCAATTGGAGAGCAATGGCTAGTGACTTTAATCTACCACCTATAGTA
          .          .          .          .          .      3900
AlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluAlaMetHisGly
GCGAAGGAAATAGTAGCCAGCTGTGATAAATGTCAACTAAAAGGGGAAGCCATGCATGGA
          .          .          .          .          .         .
GlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeuGluGlyLysIle
CAAGTAGACTGTAGTCCAGGGATATGGCAATTAGATTGCACACATCTAGAAGGAAAAATA
          .          .          .        4000       .         .
IleIleValAlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGlu
ATCATAGTAGCAGTCCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATCCCAGCAGAA
          .          .          .          .          .         .
ThrGlyGlnGluThrAlaTyrPheIleLeuLysLeuAlaGlyArgTrpProValLysVal
ACAGGACAGGAGACAGCATACTTTATACTAAAATTAGCAGGAAGATGGCCAGTAAAAGTA
          .       4100          .          .          .         .
```

```
ValHisThrAspAsnGlySerAsnPheThrSerAlaAlaValLysAlaAlaCysTrpTrp
GTACACACAGACAATGGCAGCAATTTCACCAGTGCTGCAGTTAAAGCAGCCTGTTGGTGG
        .         .         .         .         .      4200
AlaAsnIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGlnGlyValValGlu
GCAAATATCAAACAGGAATTTGGAATTCCCTACAACCCCCAAAGTCAAGGAGTAGTGGAA
        .         .         .         .         .         .
SerMetAsnLysGluLeuLysLysIleIleGlyGlnValArgGluGlnAlaGluHisLeu
TCTATGAATAAGGAATTAAAGAAAATCATAGGGCAGGTAAGAGAGCAAGCTGAACACCTT
        .         .         .      4300         .         .
LysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLysGlyGlyIleGly
AAGACAGCAGTACAAATGGCAGTGTTCATTCACAATTTTAAAAGAAAAGGGGGGATTGGG
        .         .         .         .         .         .
GlyTyrSerAlaGlyGluArgIleIleAspMetIleAlaThrAspIleGlnThrLysGlu
GGGTACAGTGCAGGGGAAAGAATAATAGACATGATAGCAACAGACATACAAACTAAAGAA
        .      4400         .         .         .         .
LeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArgAspAsnArgAsp
TTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAACAGAGAC
        .         .         .         .         .      4500
ProIleTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaValValIleGln
CCAATTTGGAAAGGACCAGCAAAACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACAG
        .         .         .         .         .  →Q     .
AspAsnSerAspIleLysValValProArgArgLysAlaLysIleIleArgAspTyrGly
                                                       MetGlu
GACAATAGTGATATAAAGGTAGTACCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGA
        .         .         .      4600  POL←        .      .
LysGlnMetAlaGlyAspAspCysValAlaGlyGlyGlnAspGluAsp
 AsnArgTrpGlnValMetIleValTrpGlnValAspArgMetArgIleArgThrTrpHis
AAACAGATGGCAGGTGATGATTGTGTGGCAGGTGGACAGGATGAGGATTAGAACATGGCA
        .         .         .         .         .         .
 SerLeuValLysHisHisMetTyrValSerLysLysAlaLysAsnTrpPheTyrArgHis
CAGTTTAGTAAAACATCATATGTATGTCTCAAAGAAAGCTAAAAATTGGTTTTATAGACA
        .      4700         .         .         .         .
 HisTyrGluSerArgHisProLysValSerSerGluValHisIleProLeuGlyAspAla
TCACTATGAAAGCAGGCATCCAAAAGTAAGTTCAGAAGTACACATCCCACTAGGGGATGC
        .         .         .         .         .      4800
 ArgLeuValValArgThrTyrTrpGlyLeuGlnThrGlyGluLysAspTrpHisLeuGly
TAGATTAGTAGTAAGAACATATTGGGGTCTGCAAACAGGAGAAAAAGACTGGCACTTGGG
        .         .         .         .         .         .
 HisGlyValSerIleGluTrpArgGlnLysArgTyrSerThrGlnLeuAspProAspLeu
TCATGGGGTCTCCATAGAATGGAGGCAGAAAAGATATAGCACACAACTAGATCCTGACCT
        .         .         .      4900         .         .
 AlaAspGlnLeuIleHisLeuTyrTyrPheAspCysPheSerGluSerAlaIleArgGln
AGCAGACCAACTGATTCATCTGTACTATTTTGATTGTTTTTCAGAATCTGCCATAAGACA
        .         .         .         .         .         .
 AlaIleLeuGlyHisIleValSerProArgCysAspTyrGlnAlaGlyHisAsnLysVal
AGCCATATTAGGACATATAGTTAGTCCTAGGTGTGATTATCAAGCAGGACATAACAAGGT
        .      5000         .         .         .         .
 GlySerLeuGlnTyrLeuAlaLeuThrAlaLeuIleAlaProLysLysThrArgProPro
AGGATCTTTACAGTATTTGGCACTAACAGCATTAATAGCACCAAAAAAGACAAGGCCACC
        .         .         .  →R     .         .      5100
                                     MetGluGlnAlaProAlaAspGlnGly
 LeuProSerValArgLysLeuThrGluAspArgTrpAsnLysProGlnGlnThrLysGly
TTTGCCTAGTGTTAGGAAGCTAACAGAAGATAGATGGAACAAGCCCCAGCAGACCAAGGG
```

```
                                      Q
ProGlnArgGluProHisAsnGluTrpThrLeuGluLeuLeuGluGluLeuLysGlnGlu
 HisArgGlySerHisThrMetAsnGlyHis
CCACAGAGGGAGCCACACAATGAATGGACATTAGAACTTTTAGAGGAGCTTAAGCAAGAA
                                          5200
AlaValArgHisPheProArgIleTrpLeuHisSerLeuGlyGlnHisIleTyrGluThr
GCTGTCAGACACTTTCCTAGGATATGGCTCCATAGTTTAGGACAACATATCTATGAAACT

TyrGlyAspThrTrpGluGlyValGluAlaIleIleArgSerLeuGlnGlnLeuLeuPhe
TATGGGGATACCTGGGAAGGAGTTGAAGCTATAATAAGAAGTCTGCAACAACTGCTGTTT
                  5300
IleHisPheArgIleGlyCysGlnHisSerArgIleGlyIleThrArgGlnArgArgAla
ATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTACTCGACAGAGAAGAGCA
                                                         5400
    S          R
ArgAsnGlySerSerArgSer
    MetAspProValAspProAsnLeuGluProTrpAsnHisProGlySerGlnProArg
AGAAATGGATCCAGTAGATCCTAACTTAGAGCCCTGGAACCATCCAGGGAGTCAGCCTAG

 ThrProCysAsnLysCysTyrCysLysLysCysCysTyrHisCysGlnMetCysPheIle
GACGCCTTGTAATAAGTGTTATTGTAAAAAGTGCTGCTATCATTGCCAAATGTGCTTCAT
                                          5500
 ThrLysGlyLeuGlyIleSerTyrGlyArgLysLysArgArgGlnArgArgArgProPro
AACGAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCC
                                          S
 GlnGlyAsnGlnAlaHisGlnAspProLeuProGluGln
TCAGGGCAATCAGGCTCATCAAGATCCTCTACCAGAGCAGTAAGTAGTATATGTAATACA
                  5600
ACCTTTAGTGATATTAGCAATAGTAGCATTAGTAGTAACGCTAATAATAGCAATAGTTGT
                                                         5700
GTGGACCATAGTATTTATAGAAATTAGGAAAATAAGAAGACAAAGGAAAATAGACAGGTT
                                          ENV
                                          MetArgValArgGluIleGlnArg
GATTGATAGAATAAGAGAAAGAGCAGAAGATAGTGGCAATGAGAGTGAGGGAGATACAGA
                                          5800
  AsnTyrGlnAsnTrpTrpArgTrpGlyMetMetLeuLeuGlyMetLeuMetThrCysSer
GGAATTATCAAAACTGGTGGAGATGGGGCATGATGCTCCTTGGGATGTTGATGACCTGTA

  IleAlaGluAspLeuTrpValThrValTyrTyrGlyValProValTrpLysGluAlaThr
GTATTGCAGAAGATTTGTGGGTTACAGTTTATTATGGGGTACCTGTGTGGAAAGAAGCAA
                  5900
  ThrThrLeuPheCysAlaSerAspAlaLysSerTyrGluThrGluValHisAsnIleTrp
CCACTACTCTATTTTGTGCATCAGATGCTAAATCATATGAAACAGAAGTACATAACATCT
                                                         6000
  AlaThrHisAlaCysValProThrAspProAsnProGlnGluIleGluLeuGluAsnVal
GGGCTACACATGCCTGTGTACCCACGGACCCCAACCCACAAGAAATAGAACTGGAAAATG

  ThrGluGlyPheAsnMetTrpLysAsnAsnMetValGluGlnMetHisGluAspIleIle
TCACAGAAGGGTTTAACATGTGGAAAAATAACATGGTGGAGCAGATGCATGAGGATATAA
                                          6100
```

```
SerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrProLeuCysValThrLeu
TCAGTTTATGGGATCAAAGCCTAAAACCATGTGTAAAGCTAACCCCACTCTGTGTCACTT

AsnCysThrAsnValAsnGlyThrAlaValAsnGlyThrAsnAlaGlySerAsnArgThr
TAAACTGCACTAATGTGAATGGGACTGCTGTGAATGGGACTAATGCTGGGAGTAATAGGA
          6200
AsnAlaGluLeuLysMetGluIleGlyGluValLysAsnCysSerPheAsnIleThrPro
CTAATGCAGAATTGAAAATGGAAATTGGAGAAGTGAAAAACTGCTCTTTCAATATAACCC
                                                        6300
ValGlySerAspLysArgGlnGluTyrAlaThrPheTyrAsnLeuAspLeuValGlnIle
CAGTAGGAAGTGATAAAAGGCAAGAATATGCAACTTTTTATAACCTTGATCTAGTACAAA

AspAspSerAspAsnSerSerTyrArgLeuIleAsnCysAsnThrSerValIleThrGln
TAGATGATAGTGATAATAGTAGTTATAGGCTAATAAATTGTAATACCTCAGTAATTACAC
                                    6400
AlaCysProLysValThrPheAspProIleProIleHisTyrCysAlaProAlaGlyPhe
AGGCTTGTCCAAAGGTAACCTTTGATCCAATTCCCATACATTATTGTGCCCCAGCTGGTT

AlaIleLeuLysCysAsnAspLysLysPheAsnGlyThrGluIleCysLysAsnValSer
TTGCAATTCTAAAGTGTAATGATAAGAAGTTCAATGGAACGGAAATATGTAAAAATGTCA
          6500
ThrValGlnCysThrHisGlyIleLysProValValSerThrGlnLeuLeuLeuAsnGly
GTACAGTACAATGTACACATGGAATTAAGCCAGTGGTGTCAACTCAACTGCTGTTAAATG
                                                        6600
SerLeuAlaGluGluGluIleMetIleArgSerGluAsnLeuThrAspAsnThrLysAsn
GCAGTCTAGCAGAAGAAGAGATAATGATTAGATCTGAAAATCTCACAGACAATACTAAAA

IleIleValGlnLeuAsnGluThrValThrIleAsnCysThrArgProGlyAsnAsnThr
ACATAATAGTACAGCTTAATGAAACTGTAACAATTAATTGTACAAGGCCTGGAAACAATA
                                    6700
ArgArgGlyIleHisPheGlyProGlyGlnAlaLeuTyrThrThrGlyIleValGlyAsp
CAAGAAGAGGGATACATTTCGGCCCAGGGCAAGCACTCTATACAACAGGGATAGTAGGAG

IleArgArgAlaTyrCysThrIleAsnGluThrGluTrpAspLysThrLeuGlnGlnVal
ATATAAGAAGAGCATATTGTACTATTAATGAAACAGAATGGGATAAAACTTTACAACAGG
          6800
AlaValLysLeuGlySerLeuLeuAsnLysThrLysIleIlePheAsnSerSerSerGly
TAGCTGTAAAACTAGGAAGCCTTCTTAACAAAACAAAAATAATTTTTAATTCATCCTCAG
                                                        6900
GlyAspProGluIleThrThrHisSerPheAsnCysArgGlyGluPhePheTyrCysAsn
GAGGGGACCCAGAAATTACAACACACAGTTTTAATTGTAGAGGGGAATTTTTCTACTGTA

ThrSerLysLeuPheAsnSerThrTrpGlnAsnAsnGlyAlaArgLeuSerAsnSerThr
ATACATCAAAACTGTTTAATAGTACATGGCAGAATAATGGTGCAAGACTAAGTAATAGCA
                                    7000
GluSerThrGlySerIleThrLeuProCysArgIleLysGlnIleIleAsnMetTrpGln
CAGAGTCAACTGGTAGTATCACACTCCCATGCAGAATAAAACAAATTATAAATATGTGGC

LysThrGlyLysAlaMetTyrAlaProProIleAlaGlyValIleAsnCysLeuSerAsn
AGAAAACAGGAAAAGCTATGTATGCCCCTCCCATCGCAGGAGTCATCAACTGTTTATCAA
          7100
IleThrGlyLeuIleLeuThrArgAspGlyGlyAsnSerSerAspAsnSerAspAsnGlu
ATATTACAGGGCTGATATTAACAAGAGATGGTGGAAATAGTAGTGACAATAGTGACAATG
                                                        7200
```

```
ThrLeuArgProGlyGlyGlyAspMetArgAspAsnTrpIleSerGluLeuTyrLysTyr
AGACCTTAAGACCTGGAGGAGGAGATATGAGGGACAATTGGATAAGTGAATTATATAAAT

LysValValArgIleGluProLeuGlyValAlaProThrLysAlaLysArgArgValVal
ATAAAGTAGTAAGAATTGAACCCCTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGG
                                        7300
GluArgGluLysArgAlaIleGlyLeuGlyAlaMetPheLeuGlyPheLeuGlyAlaAla
TGGAAAGAGAAAAAAGAGCAATAGGACTAGGAGCCATGTTCCTTGGGTTCTTGGGAGCAG

GlySerThrMetGlyAlaAlaSerLeuThrLeuThrValGlnAlaArgGlnLeuLeuSer
CAGGAAGCACGATGGGCGCAGCGTCACTAACGCTGACGGTACAGGCCAGACAGTTACTGT
              7400
GlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGlnHisLeuLeu
CTGGTATAGTGCAACAGCAAAACAATTTGCTGAGGGCTATAGAGGCGCAACAGCATCTGT
                                                        7500
GlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaValGluArgTyr
TGCAACTCACGGTCTGGGGCATTAAACAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGAT

LeuGlnAspGlnArgLeuLeuGlyMetTrpGlyCysSerGlyLysHisIleCysThrThr
ACCTACAGGATCAACGGCTCCTAGGAATGTGGGGTTGCTCTGGAAAACACATTTGCACCA
                                        7600
PheValProTrpAsnSerSerTrpSerAsnArgSerLeuAspAspIleTrpAsnAsnMet
CATTTGTGCCTTGGAACTCTAGTTGGAGTAATAGATCTCTAGATGACATTTGGAATAATA

ThrTrpMetGlnTrpGluLysGluIleSerAsnTyrThrGlyIleIleTyrAsnLeuIle
TGACCTGGATGCAGTGGGAAAAAGAAATTAGCAATTACACAGGCATAATATACAACTTAA
              7700
GluGluSerGlnIleGlnGlnGluLysAsnGluLysGluLeuLeuGluLeuAspLysTrp
TTGAAGAATCGCAAATCCAGCAAGAAAAGAATGAAAAGGAATTATTGGAATTGGACAAGT
                                                        7800
AlaSerLeuTrpAsnTrpPheSerIleSerLysTrpLeuTrpTyrIleArgIlePheIle
GGGCAAGTTTGTGGAATTGGTTTAGCATATCAAAATGGCTGTGGTATATAAGAATATTCA

IleValValGlyGlyLeuIleGlyLeuArgIleIlePheAlaValLeuSerLeuValAsn
TAATAGTAGTAGGAGGCTTAATAGGTTTAAGAATAATTTTTGCTGTGCTTTCTTTAGTAA
                                        7900
ArgValArgGlnGlyTyrSerProLeuSerLeuGlnThrLeuLeuProThrProArgGly
ATAGAGTTAGGCAGGGATACTCACCTCTGTCGTTGCAGACCCTCCTCCCAACACCGAGGG

ProProAspArgProGluGlyIleGluGluGluGlyGlyGluGlnGlyArgGlyArgSer
GACCACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGCAAGGCAGAGGCAGAT
              8000
IleArgLeuValAsnGlyPheSerAlaLeuIleTrpAspAspLeuArgAsnLeuCysLeu
CAATTCGATTGGTGAACGGATTCTCAGCACTTATCTGGGACGACCTGAGGAACCTGTGCC
                                                        8100
PheSerTyrHisArgLeuArgAspLeuLeuLeuIleAlaThrArgIleValGluLeuLeu
TCTTCAGTTACCACCGCTTGAGAGACTTACTCTTAATTGCAACGAGGATTGTGGAACTTC

GlyArgArgGlyTrpGluAlaLeuLysTyrLeuTrpAsnLeuLeuGlnTyrTrpGlyGln
TGGGACGCAGGGGGTGGGAAGCCCTCAAATATCTGTGGAATCTCCTGCAATATTGGGGTC
                                        8200
```

```
 GluLeuLysAsnSerAlaIleSerLeuLeuAsnThrThrAlaIleAlaValAlaGluCys
AGGAACTGAAGAATAGTGCTATTAGCTTGCTTAATACCACAGCAATAGCAGTAGCTGAAT

 ThrAspArgValIleGluIleGlyGlnArgPheGlyArgAlaIleLeuHisIleProArg
GCACAGATAGGGTTATAGAAATAGGACAAAGATTTGGTAGAGCTATTCTCCACATACCTA
                8300
                                            F
                                    EW     MetGlyGlyLysTrpSerLys
 ArgIleArgGlnGlyPheGluArgAlaLeuLeu
GAAGAATTAGACAGGGCTTCGAAAGGCTTTGCTATAACATGGGTGGCAAGTGGTCAAAA
                                                            8400
SerSerIleValGlyTrpProLysIleArgGluArgIleArgArgThrProProThrGlu
AGTAGCATAGTAGGATGGCCTAAGATTAGGGAAAGAATAAGACGAACTCCCCCAACAGAA

ThrGlyValGlyAlaValSerGlnAspAlaValSerGlnAspLeuAspLysCysGlyAla
ACAGGAGTAGGAGCAGTATCTCAAGATGCAGTATCTCAAGATTTAGATAAATGTGGAGCA
                                        8500
AlaAlaSerSerSerProAlaAlaAsnAsnAlaSerCysGluProProGluGluGluGlu
GCCGCAAGCAGCAGTCCAGCAGCTAATAATGCTAGTTGTGAACCACCAGAAGAAGAGGAG

GluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLysGlyAla
GAGGTAGGCTTTCCAGTCCGTCCTCAGGTACCTTTAAGACCAATGACTTATAAAGGAGCT
                8600                    U3
PheAspLeuSerHisPheLeuLysGluLysGlyGlyLeuAspGlyLeuValTrpSerPro
TTTGATCTCAGCCACTTTTTAAAAGAAAAGGGGGGACTGGATGGGTTAGTTTGGTCCCCA
                                                            8700
LysArgGlnGluIleLeuAspLeuTrpValTyrHisThrGlnGlyTyrPheProAspTrp
AAAAGACAAGAAATCCTTGATCTGTGGGTCTACCACACACAAGGCTACTTCCCTGATTGG

GlnAsnTyrThrProGlyProGlyIleArgPheProLeuThrPheGlyTrpCysPheLys
CAGAATTACACACCAGGGCCAGGGATTAGATTCCCACTGACCTTCGGATGGTGCTTTAAG
                                        8800
LeuValProMetSerProGluGluValGluGluAlaAsnGluGlyGluAsnAsnCysLeu
TTAGTACCAATGAGTCCAGAGGAAGTAGAGGAGGCCAATGAAGGAGAGAACAACTGTCTG

LeuHisProIleSerGlnHisGlyMetGluAspAlaGluArgGluValLeuLysTrpLys
TTACACCCTATTAGCCAACATGGAATGGAGGACGCAGAAAGAGAAGTGCTAAAATGGAAG
                8900
PheAspSerSerLeuAlaLeuArgHisArgAlaArgGluGlnHisProGluTyrTyrLys
TTTGACAGCAGCCTAGCACTAAGACACAGAGCCAGAGAACAACATCCGGAGTACTACAAA
 F                                                          9000
AspCys
GACTGCTGACACAGAAGTTGCTGACAGGGGACTTTCCGCTGGGGACTTTCCAGGGGAGGC

GTAACTTGGGCGGGACCGGGGAGTGGCTAACCCTCAGATGCTGCATATAAGCAGCTGCTT
      U3   R                            9100
TTCGCCTGTACTGGGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGC

AAGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAA
                9200
```

INTERNATIONAL SEARCH REPORT

International Application No PCT/EP 87/00326

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC[4]: C 12 N 15/00; C 12 N 7/00; C 12 Q 1/70; G 01 N 33/569; C 07 K 13/00; C 07 K 15/00; C 12 N 1/20; C 12 N 5/00

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
IPC[4]	C 12 N G 01 N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [*]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X	Cell, volume 40, January 1985, MIT, S. Wain-Hobson et al.: "Nucleotide sequence of the AIDS virus, LAV", pages 9-17 see figure 1 --	1-19
X	Science, volume 229, 23 August 1985, F. Wong-Staal et al.: "Genomic diversity of human T-lymphotropic virus type III (HTLV-III)", pages 759-762 see table 1 --	1-19
X	The Lancet, 23 June 1984, A. Ellrodt et al.: "Isolation of human T-lymphotropic retrovirus (LAV) from Zairian married couple, one with AIDS, one with prodromes", pages 1383-1385 see the whole document	1-19

----

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
21st September 1987	21 OCT 1987
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	M. VAN MOL